Investigations of Peptide Structural Stability *in vacuo*

by

Jason Michael Drosos Kalapothakis

Ph.D.
The University of Edinburgh
2009

(Studirzimmer.)

FAUST

Aber was muß ich sehen!
Kann das natürlich geschehen?
Ist es Schatten? ist's Wirklichkeit?
Wie wird mein Pudel lang und breit!
Er hebt sich mit Gewalt,
Das ist nicht eines Hundes Gestalt!
Welch ein Gespenst bracht' ich ins Haus!
Schon sieht er wie ein Nilpferd aus,
Mit feurigen Augen, schrecklichem Gebiß.
O! du bist mir gewiß!
[...]
Hinter den Ofen gebannt
Schwillt es wie ein Elephant,
Den ganzen Raum füllt es an,
Es will zum Nebel zerfließen.
Steige nicht zur Decke hinan!
Lege dich zu des Meisters Füßen!
Du siehst daß ich nicht vergebens drohe.
Ich versenge dich mit heiliger Lohe!
Erwarte nicht
Das dreymal glühende Licht!
Erwarte nicht
Die stärkste von meinen Künsten!

MEPHISTOPHELES

(tritt, indem der Nebel fällt, gekleidet wie ein fahrender Scholastikus, hinter dem Ofen hervor.)

Wozu der Lärm? was steht dem Herrn zu Diensten?

FAUST

Das also war des Pudels Kern!
Ein fahrender Scolast? Der Casus macht mich lachen.

(Study.)

FAUST

But what is this I now must see!
Can that happen naturally?
Is it phantom? Is it reality?
How long and broad the poodle grows!
He rises up in mighty pose,
'Tis not a dog's form that he shows!
What spectre have I sheltered thus?
He's like a hippopotamus
With fiery eyes, jaws terrible to see.
Oh, mine you are most certainly.
[...]
Behind the stove, held by my spells,
Like an elephant it swells,
And all the space it fills complete.
In vapour it will melt away.
Mount not up to the ceiling! Lay
Thyself down at thy Master's feet!
I threaten not in vain as thou canst see.
With holy fire I'll shrivel thee!
Do not await
The light thrice radiate!
Do not await
The strongest art at my command!

MEPHISTOPHELES

(steps forth from behind the stove while the vapour is vanishing. He is dressed as a travelling scholar.)

Wherefore this noise? What does my lord command?

FAUST

So this, then, was the Poodle's core!
A travelling scholar it is? The *casus* makes me laugh.

Johann Wolfgang von Goethe
Faust: 1247-1256; 1310-1324

(translation based on that by George Madison Priest)

Acknowledgements

Completion of the work described herein, in its entirety, would not have been possible without contributions of numerous colleagues, advisors, friends and relatives, to whom I am truly indebted.

I am very grateful to my family, whose prompting was the prime cause for me undertaking this work, and whose financial and emotional support allowed its completion, and to whom this thesis is dedicated.

I would like to thank all members of the PBRG, past and present, for providing such a fruitful, pleasant and unique social environment for carrying out this research. I am particularly indebted to Bryan, who first put the MoQToF together, to Hayden for all his help on simulations as well for setting up and maintaining the group's cluster, to Wutharath for help with QM calculations, scientific discussions and delicious *tartes*, to Hannah the colourful for her advice on using the LCQ, practical mass-spec and spectroscopic advice as well as her melodious singing, to cunning Pete, whose constantly cheerful demeanor and linguistic aptness was always deemed most penetrating and was almost invariably well received; to Stefan, Fiona, Martin, Roland, Yana, Jude and Holger for the invaluable assistance, discussions and coffee breaks; to all those fellow group members involved directly to the projects that I also occasionally occupied myself: Andrew Stopford, who carried out CID and HDX experiments on Trp cage, Claire McMillan, who also performed HDX and spectroscopy of Trp cage; Guy Bennet and Hattie Cole who worked very patiently on TTR. I ought to express my amazement to the patience, diligence and dedication that all of them showed towards their work. I would also like to thank members of Dr. P.J. Camp's (esp. Alex Chremos, Georg Ganzenmueller) and Prof. P.A. Madden's (esp. Lindsay Foy and Michael Pounds) groups, as well as all other residents of the old library, for useful discussions on computational chemistry and a kaleidoscope of other topics.

Numerous collaborators contributed very significantly to this work: I am very obliged to Dr. Cait MacPhee for allowing me to work on the TTR project and to Helen Cooper for offering the opportunity to work on the L_4PL_4K peptide; it was also a great joy to attend some Team Defensin meetings albeit that my research

eventually diverted from the systems examined by this group. I am also very indebted to Dr. Derek Macmillan for providing synthetic Trp cage peptides, to Prof. Lindsay Sawyer for access on the Jasco spectropolarimeter, to Dr. David Dryden for access to his fluorescence spectrometer, Dr. Philip Camp for explaining the solution of the transport equation, Dr. Bridgette Duncombe and Dr. Dušan Uhrin for their advice. Secretarial, administrative, IT and technical staff in the School of Chemistry also helped significantly for carrying out this work, so I would like to express my gratitude to all.

I would like to acknowledge all my friends and acquaintances who made my stay in Edinburgh worth living.

Finally, I would like to say that I am deeply indebted to my main suvervisor, Dr. Perdita Barran, who took me under her proverbial wing and gave me the opportunity to be involved in such an intriguing, interdisciplinary and rapidly developing area of research. Without her constant support I would not be able to complete any of this work.

Declaration

The entire *corpus* of work described herein has been performed by myself unless otherwise indicated. All other contributors to this work have been acknowledged appropriately. This thesis is submitted to the University of Edinburgh towards completion of the degree of Doctor of Philosophy and has not been submitted in whole or in part to any other institution for the acquisition of any other qualification, in accordance with University regulations.

Jason M. D. Kalapothakis

Table of Contents

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Appendices

Abbreviation List

Aβ	Alzheimer's β peptide
ACN	Acetonitrile
AFM	Atomic Force Microscopy
AMBER	Molecular Dymanics software package: Assisted Model Building with Energy Refinement
ATD	Arrival Time Distribution
BIRD	Blackbody Infrared Radiation Dissociation
BK	Bradykinin
CD	Circular Dichroism
CID	Collisionally-Induced Dissociation
DiP	Defensin-inspired-Peptide
EHSS	Exact Hard Sphere Scattering Model
EM	Electron Microscopy
ESI	Electrospray Ionisation
FAC	Familial Amyloidotic Cardiomyopathy
FAIMS	high Field Asymmetric waveform Ion Mobility Spectrometry
FAP	Familial Amyloidotic Polyneuropathy
FCS	Fluorescence Correlation Spectroscopy
FTIR	Fourier Transform InfraRed spectroscopy
HDX	Hydrogen-Deuterium eXhange
IMS	Ion Mobility Spectrometry
IMMS	Ion Mobility / Mass Spectrometry
LMW	Low Molecular Weight
MALDI	Matrix-Assisted Laser Desorption / Ionisation
MC	Monte Carlo (simulation)
MCPs	MicroChannel Plates
MD	Molecular Dynamics
MM	Molecular Mechanics
MMC	Metropolis Monte Carlo (simulation)
MoQToF	Mobility Quadrupole Time-of-Flight
MS	Mass Spectrometry
nESI	Nano-Electrospray Ionisation
pdb	Protein Data Bank
PA	Projection Approximation (of the collision cross section)
PEG	Poly(Ethylene Glycol)
PEO	Poly(Ethylene Oxide)
PES	Potential Energy Surface
QIT	Quadrupole Ion Trap
QLS	Quasi-elastic Light Scattering
QToF	Quadrupole Time-of-Flight
REM	Replica Exchange Method
RF	Radio-Frequency (alternating current)
sander	(MD propagation) Simulated Annealing with NMR-type Restraints
SEC	Size-Exclusion Chromatography
SSA	Senile Systemic Amyloidosis
TFE	Trifuoroethanol

ThT	Thioflavin T
TJ	Trajectory integration method
ToF	Time of Flight (mass analyser)
TTR	Transthyretin (equivalent to prealbumin)
TWIG	Travelling Wave Ion Guide
TWIMS	Travelling Wave Ion Mobility Spectrometry

Physical Quantities and Units

D	Diffusion coefficient
E	Energy; Electric field
$\mathbf{F}$	Force
$\mathbf{g}$	Relative velocity
h	Plank constant
I	Current or Intensity
$\mathbf{J}$	Particle flux
k_B	Boltzmann constant
K	Mobility
l	length
m	Mass
M	Molar aount; moles per litre
n,N	Number, or particle number density
O	Observable
P	Pressure
Q	Heat
q	Net charge (q ≡ ze)
R	Resistance
T	Temperature
t	Time
$\mathbf{v}_\mathrm{D}$	Drift velocity
$\mathbf{v}_\mathrm{r}$, $\mathbf{v}_\mathrm{rel}$	Relative velocity between two point masses
z	Nominal charge
x, $\mathbf{X}$, r, $\mathbf{r}$	Coordinate, position, configuration, distance ($r$adius)
(x, y, z)	Rectangular coordinates
(k, l, m)	Reciprocal space vector
θ, φ, γ	Angles
χ	Scattering Angle (of an ion-neutral collision)
Ω	Collision integral (momentum transfer cross section)

Abstract

Gas-phase analytical techniques provide very valuable tools for tackling the structural complexity of macromolecular structures such as those encountered in biological systems. Conformational dynamics of polypeptides and polypeptide assemblies underlie most biological functionalities, yet great difficulties arise when investigating such phenomena with the well-established techniques of X-ray crystallography and NMR. In areas such as these ion mobility interfaced with mass spectrometry (IMMS) and molecular modelling can make a significant contribution.

During an IMMS experiment analyte ions drift in a chamber filled with an inert gas; measurement of the transport properties of analyte ions under the influence of a weak electric field can lead to determination of the orientationally-averaged collision cross-section of all resolved ionic species. A comparison with cross-sections estimated for model molecular geometries can lead to structural assignments. Thus IMMS can be used effectively to separate gas-phase ions based on their conformation. The drift tube employed in the experiments described herein is thermally regulated, which also enables the determination of collision cross-sections over a range of temperatures, and can provide a view of temperature-dependent conformational dynamics over the experimental (low microsecond) timescale.

Studies described herein employ IMMS and a gamut of other MS-based techniques, solution spectroscopy and – importantly – molecular mechanics simulations to assess a) conformational stability of isolated peptide ions, with a focus on small model peptides and proteins, especially the Trp cage miniprotein; and b) structural characteristics of oligomeric aggregates of an amyloidogenic peptide.

The results obtained serve to clarify the factors which dominate the intrinsic stability of non-covalent structure in isolated peptides and peptide assemblies. Strong electrostatic interactions are found to play a pivotal role in determining the conformations of isolated proteins. Secondary structures held together by hydrogen bonding, such as helices, are stable in the absence of solvent, however gas-phase protein structures display loss of their hydrophobic cores. The absence of a polar solvent, “self-solvation” is by far the most potent force influencing the gas-phase

configuration of these systems. Geometries that are more compact than the folded state observed in solution are routinely detected, indicating the existence of intrinsically stable compact non-native states in globular proteins, illuminating the nature of proteins' 'unfolded' states.

1

Structural investigation of peptides by means of a combined theoretical/experimental approach

1.1 Prologue

Recent decades have witnessed a revolution in the speed of DNA sequencing processes resulting to a steady increase in the number of complete genome sequences becoming available. Such developments create an urgent need for rapid, reliable and cost-effective methods that can analyse this vast amount of sequence data. Therefore it is not surprising that the assignment of biological function(s) to amino acid sequences constitutes a major challenge for biological sciences.

One may intuitively state that the ability to describe biological processes at a molecular level alone can achieve the greatest possible understanding of biological systems. So far, molecular structures of biopolymers have been obtained mostly by experimental techniques such as X-ray crystallography and NMR. Unfortunately these extremely powerful methods often fail to adequately describe flexible structures or systems in which numerous different geometries coexist. Now, considerable effort has been –and is being– directed towards obtaining structural information from an amino-acid sequence computationally. Homology modelling methods attempt to fit a peptide sequence to that of a homologous protein for which a structure has been solved experimentally. Such an approach can be justified by the finding that proteins with even less than 30% sequence similarity adopt highly similar folds[2] (Figure 1.1). Nevertheless such an approach presents considerable drawbacks:

- The importance of structural flexibility is – once again – overlooked

- Misleading results may be obtained due to errors in the experimentally determined structure
- A significant number of putative amino acid sequences is being obtained for which there is no sequence homology to peptides with "known" structures.

Given the timescale over which folding of an average-sized/large 'globular' protein (many hundreds of amino acids in length) takes place, milliseconds to minutes, exhaustive simulation of folding 'pathways' by molecular mechanics calculations is not a viable strategy for currently available computational resources. Nevertheless, computational efficiency has been increasing exponentially for several decades (Figure 1.2[3]) and this trend is expected to continue for the immediate future; thus an era when protein structure can be predicted from sequence is not entirely beyond all realistic aspiration. The structures of smaller peptides and "miniproteins" (several tens of amino acids in length) can be represented computationally at present time to such a level of precision that achieves reasonably good agreement with experimental results. The ability to probe peptide structure computationally opens a number of new intriguing directions for biochemical and biophysical research. Smaller peptides often have poorly defined structures which are heavily dependent on environmental conditions. Indeed, such heterogeneity of environments that are encountered *in vivo* is typical of such systems. Peptide-protein and peptide-membrane interactions are essential for a wide range of physiological phenomena, e.g. cell signalling, membrane turnover and – ultimately – protein folding, which encompass large-scale structural transitions and cannot be adequately explored at an atomic resolution by focusing on very few stable structures. Therefore a very wide range of biochemical phenomena ask for a way by which structural flexibility and dynamics can be accounted for.

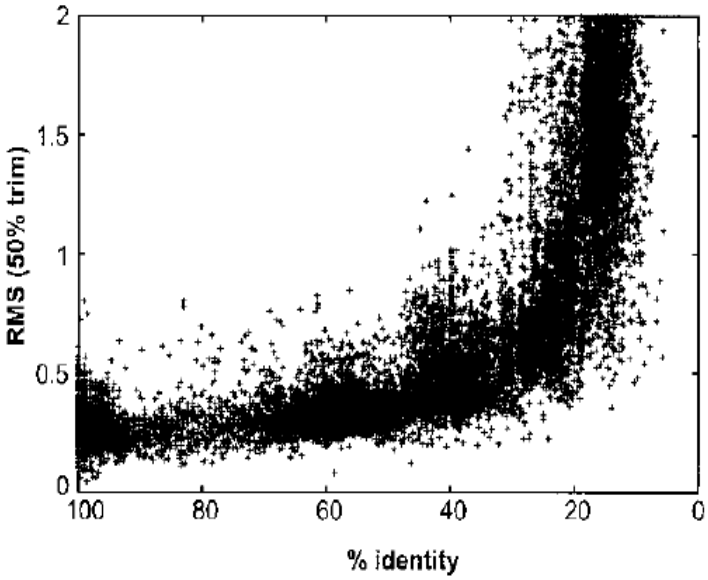


Figure 1.1
Sequence identity vs. backbone RMSD for a large set of protein crystal structures. Reproduced from ([2])

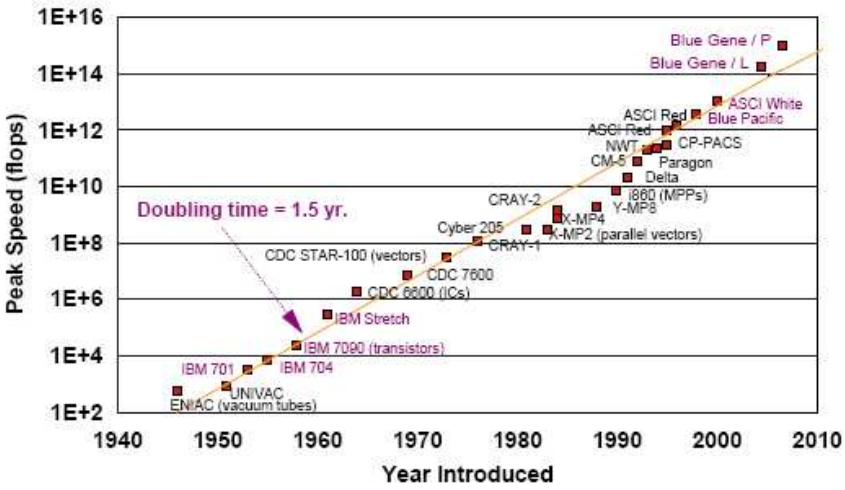


Figure 1.2
Exponential increase in processor power
Source: http://www.research.ibm.com/bluegene

1.1.1 Motivation for Research

Peptide structure is traditionally explored spectroscopically in dilute aqueous solution or in the solid phase by x-ray scattering techniques. The gas phase presents an additional environment in which peptides can be isolated and analysed. Due to the alien nature of such a medium in biological research the relevance of such endeavours can be questioned. Thankfully, a simple rhetoric can dispel such scepticism. Studies of peptide structure in the absence of other strongly interacting partners can provide a picture of the "intrinsic" structural tendencies of the molecules of interest. Formations that arise directly from the chemical composition of a particular molecule may be studied in isolation. Solvent effects can be looked into as well, by micro-solvation experiments. An aqueous environment is not necessarily the 'best' medium for studying biological systems: *e.g.* membrane proteins, which contribute significantly to any organism's proteome, encounter environments with dielectric constants much closer to that of a vacuum than that of water. Choosing the gas phase, or a vacuum, as a medium also presents numerous practical advantages. A wide range of analytical techniques, notably mass spectrometry (MS) and MS-based methods, exist, and can be employed to address problems of protein conformation. A vacuum is a medium that can be represented straightforwardly by computational means as well, since the representation of solvent effects by empirical means is no simple task, especially when the solvent in question is water. The gas phase constitutes a relatively unexplored medium for biomolecule studies yet it evidently deserves the attention of the broader research community.

Herein results on the conformation and structural flexibility of two model peptide system are presented. The principal experimental techniques employed are ion mobility spectrometry (IMS) integrated with MS. An effort to approach the problems in question from a theoretical (or more aptly: computational perspective) is also made. Occasionally other techniques (both in the gas and condensed phases) are also employed. The systems in question are Trp cage – the smallest peptide known at the time of the beginning of this project which exhibits significant similarities to larger well-folded globular proteins – and human Transthyretin (TTR) fragment 105-115, a peptide that spontaneously self assembles to form amyloid fibrils *in vitro*. Trp cage is

a model system for investigating the folding of monomeric proteins and their structural stability in vacuo, whereas TTR105-115 is a model system used to explore peptide self-assembly. IMMS was chosen as the experimental point of focus due to its analytical strengths, in particular its ability to separate different species, including different conformations, from their transport properties in an inert gas.
The structure of the thesis may now be summarised.

1.1.2 Overview of Thesis

Chapter 1 (this chapter) contains a summary of the physical principles of the techniques used in our research strategy. Computational methods are described first. The discussion begins with describing the usefulness of empirical potentials for classifying molecular geometries. The problems encountered when studying macromolecular systems are then discussed. The possibility of following the dynamical behaviour by employing these empirical methods in conjunction with classical mechanics is also explained. The principles of sampling over the entire set of possible molecular (nuclear) geometries are also briefly exposed. Following that, the link between transport properties and geometry is shown, forming the basis for the usefulness of IMMS in addressing problems of biomolecular structure. The quantities of interest are defined before the argument goes on into a review of the application of IMMS on questions of protein conformation. Some general traits of instrumentation are also exposed. Finally, the chapter closes with a brief description of other MS-based techniques relevant to the results presented herein.
Chapter 2 consists of practical details concerning the implementation of the techniques in the strategy that is described in Chapter 1. Specific protocols are described as well as their applicability and shortcomings. To retain analogy with Chapter 1, simulation methods are described first, followed by MS and IMMS-related instrumentation and protocols. MS-based techniques described also include collision-induced dissociation (CID) and gas-phase hydrogen-deuterium exchange (HDX). Finally, the usage of circular dichroism (CD) spectroscopy is also described.
In Chapter 3 results on Trp cage ions obtained by IMMS at room temperature (T≈300K), CID, HDX, CD, and simulations including energy optimisation by simulated annealing and molecular dynamics simulations (MD) are discussed. Arg-Lys variants of Trp cage, both with ‘free’ and protected termini, are synthesised to

address the effect of charge location on peptide structure; a construct containing a *d*-amino acid is also described, to elucidate the overall structural characteristics of a peptide analogue expected to be unfolded in solution.

Chapter 4 contains results obtained by IMMS at elevated temperatures, supplemented by high-temperature MD and CD spectroscopy. This chapter addresses the thermal stability of peptide structure *in vacuo*, using Trp cage (and its variants) as a model.

In Chapter 5 results obtained on the amyloidogenic endekapeptide TTR105-115 will be discussed. The results are contextualised by a relatively long review on the self-assembly of peptides into amyloid.

Chapter 6 is a general summary of the entire thesis, drawing the main conclusions of this research as a whole and putting these conclusions into a broader context.

Supplementary information is included in the appendices.

The introduction will now proceed with the description of a way to represent and classify structures of biological molecules (peptides in particular) computationally.

1.2 Representation of Biomolecular Transconformation

1.2.1 Energy Landscapes

Envisaging phase space (for a system containing N atoms, phase space corresponds to the 6N-dimensional space of all possible atomic positions and momenta) as a function of a parameter related to the probability of occupation of every microstate, *i.e.* energy, defines a multidimensional surface. A simplistic view of this surface in three dimensions would reveal recognisable topological features: mountaintops (low-probability maxima), valleys, rivers (lowest energy paths connecting high-energy and low-energy states) and lakes at the bottom of valleys (most populated, low energy states). The overall 'appearance' of the energy landscape can provide valuable information about system thermodynamics and kinetics. *E.g.*, a well-folding globular protein would have a funnel-like energy landscape whereas this surface would have a more rugged appearance for 'disordered' polymers. An algorithm that explores the conformational energy landscape can ultimately reveal the structural properties of any interesting system. Yet the complexity of such a task is realised once one is reminded of the dimensionality of this energy surface. Since "phase space" signifies the total number of states available to our system, all contributions to the system's energy should be included. Therefore a number of approximations are necessary to render biomolecular systems computationally tractable. Protein transconformational phenomena occur at a timescale ranging from microseconds to several hours; therefore one can assume that the system under study will be dominated by ground-state electronic behaviour. Also, when observing isolated biopolymers in a temperature-regulated environment, translational and rotational motion of the molecule as a whole does not provide any useful information. Biomolecular structures can be described in terms of nuclear positions and momenta alone. Thus,

$$E_{tot} = E_k(\mathbf{p}) + V(\mathbf{X}) \qquad \{1.1\}$$

where E_{tot} is the total energy of the system, $E_k(\mathbf{p})$ the kinetic energy and $V(\mathbf{X})$ the potential energy, $\mathbf{p}$ being the momenta and $\mathbf{X}$ the coordinates for each atom in our system. Using this relation, equation {1.1} can be re-written as

$$P_\alpha = \frac{e^{-\frac{E_{k_\alpha}(\mathbf{p})+V_\alpha(\mathbf{X})}{k_BT}}}{\iint e^{-\frac{E_k(\mathbf{p})+V(\mathbf{X})}{k_BT}} d\mathbf{p}d\mathbf{X}} = \frac{e^{-\frac{E_{k_\alpha}(\mathbf{p})}{k_BT}}}{\int e^{-\frac{E_k(\mathbf{p})}{k_BT}} d\mathbf{p}} \cdot \frac{e^{-\frac{V_\alpha(\mathbf{X})}{k_BT}}}{\int e^{-\frac{V(\mathbf{X})}{k_BT}} d\mathbf{X}} = P(\mathbf{p}_\alpha)P(\mathbf{X}_\alpha) \qquad \{1.2\}$$

allowing us to sample through configurational and kinetic terms separately should we wish to. Now, for a classical mechanical system the kinetic energy is equal to

$$E_k = \sum_{i=1}^{N} \frac{m_i v_i^2}{2} = \sum_{i=1}^{N} \frac{p_i^2}{2m_i} \qquad \{1.3\}$$

N being the total number of atoms and p_i, v_i, m_i the momentum, speed and mass of atom i, respectively. The momentum integral can be solved analytically allowing for the partition function $Z(\mathbf{p}, \mathbf{X})$ to be formulated as follows:

$$Z = \frac{1}{h^{3N}N!}\int e^{-\frac{\sum_{i=1}^{N}\frac{p_i^2}{2m_i}}{k_BT}} d\mathbf{p}\int e^{-\frac{V(\mathbf{X})}{k_BT}} d\mathbf{X} = \frac{1}{N!}\prod_{i=1}^{N}\left(\frac{2\pi k_B T m_i}{h^2}\right)^{\frac{3}{2}}\int e^{-\frac{V(\mathbf{X})}{k_BT}} d\mathbf{X} = Z(\mathbf{p})Z(\mathbf{X}) \qquad \{1.4\}$$

$Z(\mathbf{p})$ corresponding to the momentum partition function, $Z(\mathbf{X})$ to the configurational partition function, h to Planck's constant, the factor $1/h^{3N}N!$ being necessary for conformity with the analogous quantum mechanical expression for N identical atoms. A valuable observation resulting from this rendering is that the only system-dependent variables that have a bearing on the momentum integral are the number and masses of atoms in our system and the temperature. As such, it does not inform us about the energetic relationship between different atoms; thus the configuration integral, which is associated to a potential energy surface (PES), proves to be the most "interesting" from a structural point of view. Unfortunately, for large systems, this integral is the most daunting to compute.

1.2.2 Dimensions of Configurational Space and the Chimaera of Exhaustive Sampling ("Levinthal's Paradox")[4,5]

The dimensionality of configurational space is 3N (N = number of atoms) in Cartesian space (three coordinates for each atom), which is reduced to 3N-6 by the use of internal coordinates (since global rotation and translation do not effect the potential energy). Further simplifications can be called upon to reduce the effective size of configuration space. An example: assume all atoms being constrained to 'equilibrium' bond lengths and angles, only allowing torsions around single bonds involving heavy atoms to vary, thus reducing the degrees of freedom available to an

amino acid to 4.15, on average. Therefore, a –relatively small– protein of 100 amino acids will have a 414-dimensional configuration space[*]. Sampling of 5 points over each degree of freedom would require 5^{414} sets of computations to be made. This corresponds to at least $\sim 7.49 \cdot 10^{266}$ years of computation time for a futuristic petaFLOP/S supercomputer – i.e. many orders of magnitude more than the estimated age of the universe. For this reason, one must resort to more sophisticated numerical sampling techniques for surveying conformational space. An overview of such methods will be given in section 1.3.2. This situation is similar to the well-known Levinthal paradox, only that the protein scientist, or, to be exact, his computer is the one who samples randomly over all accessible states, and not the protein, the latter being in need of surprisingly few computational resources to fold efficiently.

1.2.3 The Potential Function

A certain degree of coarse-graining is necessary in order to render systems as large as biopolymers computationally tractable. As a result, such systems cannot be treated quantum-mechanically, therefore relying on empirical parameters that can accurately reproduce the behaviour of a broad range of chemical systems. Furthermore, it is clear from the above that an expression for evaluating the potential energy for each given state is necessary for assessing its stability. Such considerations led to the development of the Class I potential energy function early in the 1980s, which served as a basis for most atomistic empirical models since[6]:

$$V(\mathbf{X}) = \sum_{bonds} \mathcal{K}_b (b - b_{eq})^2 + \sum_{angles} \mathcal{K}_\theta (\theta - \theta_{eq})^2 + \sum_{dihedral} \frac{\mathcal{V}_n}{2}[1 + \cos(n\varphi - \gamma)] + \sum_{\substack{non-bonded\ interactions \\ i>j}} \left[\frac{\mathcal{A}_{ij}}{r_{ij}^{12}} - \frac{\mathcal{B}_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r r_{ij}} \right] \quad \{1.5\}$$

The potential contains four distinct energy terms: simple harmonic potentials are applied to adjust bond lengths (b) and angles (θ) close to their equilibrium positions (b_{eq} and θ_{eq}); a periodic function is employed to account for 1-4 interactions (φ is the torsional angle); the non-bonded interactions are represented by two pair-wise potentials dependent on the distance R_{ij} of any two atoms i and j: Coulomb's law for electrostatic interactions (q is the charge and ε_r is the dielectric constant of the medium – in a vacuum $\varepsilon_0 = 1$) and a Lennard-Jones potential for "weak"

[*] (4.15·100)-1; the N-terminal φ angle would not be considered under this scheme.

(alternatively "van der Waals") interactions. $\mathcal{K}_r$, $\mathcal{K}_\theta$, $\mathcal{V}_n$ and even the Lennard-Jones parameters $\mathcal{A}_{ij}$ and $\mathcal{B}_{ij}$ are "force constants": scaling factors that are adjusted so that the predicted results fit high-level quantum calculations and/or experimental results for small test systems.

1.2.4 Force-Field Parameterisation – Charge Derivation

Throughout this work software from the AMBER 8 package have been used for simulations. As already mentioned, force field parameters are usually derived from and adjusted to fit experimental data (often X-ray data); however, several parameters are difficult determine experimentally – including the partial charges for atoms. Now, the potential evaluation in empirical force-fields such as AMBER is dominated by electrostatics, especially when considering large systems containing several polar groups. An accurate charge derivation methodology is therefore crucial for obtaining reliable results.

Initially, parameterisation for the AMBER force field (and indeed many other widely used force fields) was based on gas-phase simulations. As more powerful processors became available, attention was driven towards simulating organic and bioorganic compounds in their condensed phases. This development is directly reflected in the evolution of charge derivation strategies employed by AMBER developers. Whereas the electrostatic potential was calculated at a STO-3G level at the Weiner *et al.* 1984 force field[7,8], the Cornell *et al.* force field[9] employed the Hartree-Fock method and the 6-31G* basis set. This constituted a major improvement at the time, yet it quickly became apparent that this method over-estimates the dipole for polar groups if compared to the "gas phase", "average" dipole. However this discrepancy was welcomed, since it was found to mimic the solvent polarisation effect that takes place in the condensed phases.

Finally further manifestations of the attempt to take into account solvent effects in charge derivation become apparent in the development of the Duan *et al.* 2003[10] parameter set, in which the ESP was derived by a B3LYP/cc-pVTZ//HF/6-31G** method, performing the quantum calculations in a solvent continuum.

1.3 Exploring (Bio)molecular Structures

So far we have described a method to represent a molecular system computationally and evaluate its potential energy by considering bond lengths, angles, torsions, van der Waals and electrostatic interactions. Sampling over configurational space is a necessary (and computationally most expensive) step. Molecular dynamics (MD; alternatively molecular mechanics, MM)-based approaches are suitable for capturing the conformational dynamics of a very narrow range of conformational ensembles and can be conceived as "real time" simulations of the system under study. Several algorithms – many of which incorporate MD albeit in a broader context – have been developed to achieve a much more extensive sampling of conformation space and will be discussed in section 1.3.2.

1.3.1 Molecular Dynamics/Mechanics

MD approaches for probing conformational changes represent each atom as a Newtonian particle and rely on Newtonian mechanics for simulating molecular motion. It is known that the force (**F**) has a gradient relationship to the internal energy (U):

$$\mathbf{F} = -\nabla U \qquad \{1.6\}$$

Furthermore,

$$\mathbf{F} = m\mathbf{a} \equiv m\frac{\partial^2 \mathbf{x}}{\partial t^2} \qquad \{1.7\}$$

Thus, given the positions and velocities (**v**) of all atoms at time t, the positions (**x**) at time $t+\tau$ (where τ is the time-step) can be estimated, e.g. using the Verlet algorithm:

$$\mathbf{x}(t+\tau) = 2\mathbf{x}(t) - \mathbf{x}(t-\tau) + \frac{\mathbf{F}}{m}\tau^2 \qquad \{1.8\}$$

$$\mathbf{v}(t) = \frac{\mathbf{x}(t+\tau) - \mathbf{x}(t-\tau)}{2\tau} \qquad \{1.9\}$$

which is accurate assuming constant acceleration. Variants of the Verlet propagator may also be useful in carrying out this task, such as e.g. the leapfrog propagator, which calculates the velocities at half-step times:

$$\mathbf{v}\left(t+\frac{\tau}{2}\right) = \mathbf{v}\left(t-\frac{\tau}{2}\right) + \frac{\mathbf{F}}{m}\tau \qquad \{1.10\}$$

$$\mathbf{x}(t+\tau) = \mathbf{x}(t) + \mathbf{v}\left(t + \frac{\tau}{2}\right)\tau \qquad \{1.11\}$$

or even the velocity Verlet algorithm, which estimates the velocities at half-times to obtain the coordinates at $t + \tau$ and then re-evaluates the velocities at t:

$$\mathbf{x}(t+\tau) = \mathbf{x}(t) + \mathbf{v}(t)\tau + \frac{\mathbf{F}(t)\tau^2}{2m} \qquad \{1.12\}$$

$$\mathbf{v}(t+\tau) = \mathbf{v}(t) + \frac{\tau\left[\mathbf{F}(t) + \mathbf{F}(t+\tau)\right]}{2m} \qquad \{1.13\}$$

The forces for each $\mathbf{x}(t+\tau)$ are derived from the energy (equation {1.5}), thus allowing the cycle to continue. Since the forces acting on each atom depend on the coordinates (i.e. mostly its neighbouring atoms) a time step τ must be chosen that will not generate very "unnatural" geometries (in which case the energy would be very high and would continue to increase at each time step). Thus the time-step must not exceed the time scale of the bond and angle vibrations for the smallest atoms (i.e. hydrogens); thus one is restricted to using a femtosecond (or a fraction of a femtosecond) time-step unless the vibrational motion is otherwise restricted. The choice of time-step though is also dependent on temperature (higher temperatures will require smaller time-steps).

If it is necessary to assign velocities at the first step, the preferred method would be random assignment of the velocity vectors at atom centres according to a Boltzmann distribution, fulfilling the following criterion:

$$\left\langle E_k \right\rangle = \frac{N_{DoF} k_B T}{2} \qquad \{1.14\}$$

where $\left\langle E_k \right\rangle$ is the average kinetic energy and N_{DoF} the number of degrees of freedom ("dimensionality") of the system. Yet classical mechanics dictates

$$E_k(\mathbf{v}) = \frac{m\mathbf{v}^2}{2} \qquad \{1.15\}$$

thus drawing a direct relationship between the velocities of the atoms in the system under study and the temperature. This relation can be employed to scale the atomic velocities in order to keep the system at a desired temperature and is one of the simplest temperature regulation methods.

Protein folding phenomena and other large-scale transconformational events take place at a time-scale of milliseconds to seconds, or even many hours; protein

dynamics can only be probed for multi-nanosecond timescales (which corresponds to many millions of calculations of coordinates and velocities), thus limiting the scope of straightforward MD simulations. The system is unlikely to locate the global energy minimum over such timescales. At high temperatures the system is going to cover larger sections of conformation space, yet stable, compact structures are unlikely to be visited. Thus a simple MD run is likely to be kinetically 'trapped' at a local 'attraction basin'. This situation creates the need for more sophisticated conformational sampling methods to be developed.

1.3.2 Conformational Sampling Methods

As discussed above, use of computational methods may be necessary for providing "high resolution" data for any of biomolecular system for which one or two geometries are insufficient for understanding its action, especially those that are structurally more heterogeneous at equilibrium, i.e. have a number of low-energy structures. Although it might be desirable to obtain the density of each of all accessible states at equilibrium by solving {1.4}, the sheer number of such states is prohibitive. On the other hand though, the same expression indicates that high-energy states will have a negligible contribution to the overall behaviour of the system (since their population is minute compared to that of lower-energy states). Thus a reasonable estimate for the configuration integral – and, consequently, system properties can be obtained if the low energy region of the system of interest is sufficiently sampled.

A random search of coordinates will mostly generate unfavourable structures (with numerous atom overlaps and unfavourable contacts). A way to bias the search of minima towards low energy structures is provided by the Metropolis criterion. During a simulation consisting of generating a new structure (**X'**) from the previous geometry ($\mathbf{X}_n$) by introducing a structural permutation, the new structure will be accepted with the following probability:

$$P_{\mathbf{X}'\rightarrow\mathbf{X}_{n+1}} = \begin{cases} 1 & \text{iff } V(\mathbf{X}') \le V(\mathbf{X}) \\ e^{-\left(\frac{V(\mathbf{X}')-V(\mathbf{X})}{k_B T}\right)} & \text{iff } V(\mathbf{X}') > V(\mathbf{X}) \end{cases} \qquad \{1.16\}$$

The Metropolis criterion is implemented computationally by means of a random number generator: if $V(\mathbf{X}')>V(\mathbf{X})$, the configuration will be accepted if a randomly generated number in the [0,1] range is smaller or equal to $e^{-\frac{\Delta V}{k_B T}}$. This expression only leaves the temperature T to be adjusted; values for T are chosen to obtain efficient exchange rates during a simulation.

The Metropolis criterion is a common feature of numerous global optimisation and conformational searching algorithms, some of which are now going to be speedily summarised.

Global optimisation strategies survey the PES in search of the global minimum, which for a system with a funnel-like PES topology will be its most characteristic conformation. As a large number of different algorithms that have been developed over time for surveying configuration space[11-14]. It would be beyond our purposes to discuss each in detail here. Instead we will focus on simulated annealing, the method used extensively for the work described herein.

Simulated annealing (SA)[15] is a widely applicable, easy to implement global optimisation technique. During an SA run, the system is heated and equilibrated at a high temperature (so that it can "diffuse" freely in conformational space) and is gradually cooled down, usually to 0K, relying on the assumption that the system will locate the lowest-energy structure during this cooling time. As such, it suffers from a number of problems, which limit its applicability to small, "easy" systems. For example, the global free energy minimum at a high T may not coincide with this at 0K; yet at a lower temperature the system may not possess sufficient energy to overcome the barriers that separate it from a more favourable region of the PES. Such behaviour would result in the system repeatedly visiting a configuration that is not the lowest-energy structure. Larger systems, which are associated with more complex potential energy landscapes, increase the difficulty of locating the global minimum since as the energy landscape becomes rougher at lower temperatures the system can easily become trapped at a local minimum if it does not occupy the lowest-energy structure at any given time. Of course for the latter to be true, a significant amount of computational time must be provided to ensure that the system is at equilibrium at any given temperature and given the size of configuration space for macromolecules, this time becomes prohibitive. Nevertheless SA has been found

to perform sufficiently well for small systems (a more detailed discussion on the convergence properties of the algorithm is given in Chapter 2) and its conceptual simplicity and ease of implementation lead to its continued use. Due to these practical advantages, SA has been adopted as the main technique employed here for conformational sampling of gas-phase peptides.

Global optimisation schemes that "bias" the search towards low-energy regions of the PES, such as the Monte-Carlo and minimisation technique[16], are more effective when dealing with larger, more challenging systems. A number of interesting strategies have been developed for dealing with the problems of the 'straightforward' MC method (such as taboo search[17], whereby the system is discouraged to re-visit regions of conformation space that have been already sampled); these and many others have been described elsewhere[18,19].

1.4 On Mapping and Analysis of Multidimensional Energy Surfaces

Describing the PES is not a trivial matter. It is the dimensionality of the energy surface that renders this task a difficult one; configurational space is 3N (or 3N-6)-dimensional yet no more than three dimensions can be visualised at one time. Again one can distinguish between two different directions when undertaking this task:

- Projecting the 3N-6 – dimensional PES onto a low-dimensional subspace
- Performing a topological analysis.

Principal Coordinates and Principal Components Analysis (both are referred to as PCA)[20,21] are an example of the former, i.e. presenting the energy landscape on a 2- or 3-dimensional surface. The coordinates of this reduced "principal" plane are chosen (and manipulated) so that most of the relational information is included.

Simpler approaches for low-dimensional representation of conformation space involve the use of variables that are easily calculated and are thought to probe the overall structural characteristics, such as the root mean square deviation (RMSD) from a reference structure, or the radius of gyration; the use of an experimentally measurable property that is a function of the nuclear coordinates may be an attractive approach.

Topological representations of the PES focus on the portrayal of the energetic relationships between sampled minima. Knowledge of both the minima and the

transition states connecting those minima is necessary for the construction of a topological map. Disconnectivity graphs[1] are an interesting example of such a representation. Disconnectivity graphs are "tree diagrams" that describe the PES topology; a schematic of a 2-dimensional PES and its corresponding disconnectivity graph is shown in Figure 1.3.

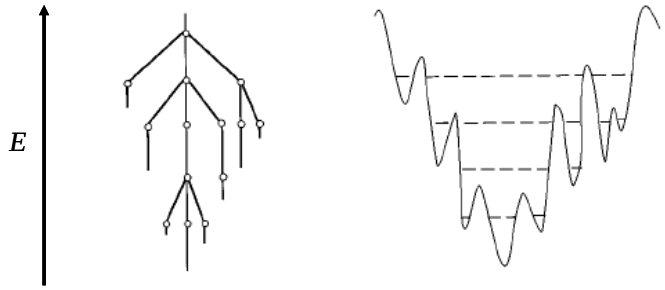


Figure 1.3
An 'energy landscape' schematic (right) accompanied by its corresponding disconnectivity graph (left).
Reproduced from *Becker & Karplus*[1]

Becker and Karplus[1] have constructed disconnectivity graphs for a capped trialanine peptide and then proceeded with solving the master equation numerically to obtain kinetic data for conformation changes in this system. Similar methods have also been reported in order to follow biomolecular dynamics from computation[1,22,23].

Simulation techniques are of great significance when the structures of studied systems can only be accessed in a low-dimensional space (as is indeed the case with ion mobility experiments). Calculation of macroscopic quantities will enable comparison with experimental data and thus give more credence to the simulation data itself. The most obvious and natural state to simulate is equilibrium (under the pre-defined parameters). For MD and MMC simulations, which satisfy detailed balance for Boltzmann weights[11,12], thermodynamic averages are obtained by averaging over all visited states so that for a quantity *O*,

$$\langle O \rangle = \overline{O} \qquad \{1.17\}$$

where $\langle O \rangle$ represents the ensemble ('thermodynamic') average and $\overline{O}$ a statistical average. However, for this relation to be true, the simulation must not only satisfy detailed balance but also be ergodic – or at least of sufficient length so that the simulated ensemble is truly *representative*. With increasingly faster computations being achievable (section 1.1) oligopeptides, despite being endowed with a rugged PES, can be treated in great detail. One may also assert that peptides of intermediate size for which an experimental structure is available a MD simulation may well be

able to locate the most populated structures in a different environment (this strategy is used for generating gas-phase structures of Trp cage and TTR as discussed in Chapters 3, 4 and 5). Non-ergodicity thus poses a problem for conformational sampling techniques; copious amounts of research have been devoted to amending this situation. Non-Boltzmann sampling has been shown to have a potential for improved sampling and is the main characteristic of generalised ensembles[24-26]. A Boltzmann probability distribution at temperature T can be calculated from a random sample of configurations (given that the sample is large enough to characterise the system over the range of desired temperatures):

$$P_B(\mathbf{X}) = \frac{e^{-\frac{E(\mathbf{X})}{k_B T}}}{\sum_{\mathbf{X}} e^{-\frac{E(\mathbf{X})}{k_B T}}} \qquad \{1.18\}$$

and $\langle O \rangle$ can be written

$$\langle O(\mathbf{X}) \rangle = \frac{\sum_{\mathbf{X}} O(\mathbf{X}) e^{-\frac{E(\mathbf{X})}{k_B T}}}{\sum_{\mathbf{X}} e^{-\frac{E(\mathbf{X})}{k_B T}}} \qquad \{1.19\}$$

Now, lets envisage a different sampling algorithm whereby a configuration X has the modified (i.e. non-Boltzmann) probability $P^*(\mathbf{X})$. To obtain Boltzmann weights the simulated probability distribution needs to be rescaled:

$$P_B(\mathbf{X}) = \frac{\left(P^*(\mathbf{X})\right)^{-1} e^{-\frac{E(\mathbf{X})}{k_B T}}}{\sum_{\mathbf{X}} \left(P^*(\mathbf{X})\right)^{-1} e^{-\frac{E(\mathbf{X})}{k_B T}}} \quad \{1.20\}$$

The estimation of $\langle O \rangle$ must be likewise modified. It is important to stress that such rescaling assumes that the sample obtained is sufficient for describing the system at the desired temperature. In practice, all sampled states are binned according to energy to obtain an energy histogram (a numerical density of states of the simulated ensemble) which is subsequently rescaled to match the expected energy distribution at a specific temperature. Equations {1.18-1.20} have been applied in cases of non-Boltzmann sampling (see, for example Hansmann & Okamoto[27]) and so we propose that they can also serve as the basis for calculation of macroscopic quantities from a

collection of simulated annealing runs, a problem alluded to in section 2.2.2.4. Nevertheless, the result needs to be stoically considered since, due to the nature of the simulated annealing scheme (described above), the rescaling's two implicit assumptions, namely

- Each energy basin is sufficiently described by its associated energy minimum
- The energy of that minimum is sufficient to describe the energetic relationship of the two states even at an elevated temperature

may not hold. Thus using experimental cross sections as "filters" for simulated geometries constitutes an alternative for a direct comparison which lacks the problems raised by temperature rescaling.

Replica exchange is remarkable in that although it allows the system to drift freely in temperature space, each constant-T replica is kept at equilibrium[28]. Thus thermodynamic data at each simulated temperature can be calculated statistically and the entire scheme can be conceived as a number of parallel hybrid MD/MC runs.

1.5 Background Information on the Experimental Methodology and Experiment-Simulation Comparison

1.5.1 The Need for Experimental Validation

Proteins are too large to be investigated by *ab initio* methods, as demonstrated in section 1.2.3, rendering the use of parameterised empirical methods necessary. Two strategies for assessing the reliability of the parameterised potentials can be easily distinguished. One is the ability of the parameterised force field to accurately mirror the energetic relationships – revealed by high-level quantum mechanical calculations – between different conformers of small test systems. This approach has become prominent in force field development[6]. The other is the ability to predict experimental data. Given the complexity of macromolecules, experimental observables constitute a stringent test for results obtained by simulation methods; also the availability of "theoretical" data can provide insights on the structure of complex systems that cannot be inferred from experiment alone (these limitations are immediate consequences of reducing the system to a low dimensionality – since the majority of experimental methods are sensitive to variation of a single degree of freedom). Such simple considerations can render a combined theoretical-experimental strategy a favourable option.

1.5.2 In Search of an Appropriate Variable

For an experimental observable to be useful in the context of a combined experimental/theoretical approach it must fulfil the following criteria:

- Sensitivity to conformational fluctuations of the system under study
- Existence of a simple method of theoretical estimation.

In this study we shall use the collision cross-section as a probe of biomolecular conformation. The collision cross-section influences the speed with which an ionised molecule drifts though a buffer gas (see section 1.5.4) and can be determined experimentally by ion mobility spectrometry (IMS). A justification of choosing a gas-phase technique for examining biomolecular structures is given in the following sections.

1.5.3 The Vapour Phase as a Medium for Studying Biomolecular Structure

Macromolecules free of a polar solvent such as water are associated with a kaleidoscope of distinct interesting phenomena, many of which are relevant to biopolymers. As mentioned above, a vacuum has a very low relative dielectric constant (ε=1) compared to, say, water at room temperature (ε=78). Short-range electrostatic interactions are much stronger in the absence of "competing" solvent molecules; furthermore the energy barriers for transconformational phenomena are heightened since long(er)-range coulombic repulsion becomes significant as it is not screened by the solvent. Consequently, multiply-charged molecules have been found to unfold and adopt extended configurations at higher charge states[29-34]. The absence of solvent destabilises charged moieties: zwitterions and salt-bridges are encountered less often in the gas phase than in solution. In the absence of system-specific conformational constraints (as in the proline-rich bradykinin[35-38]), ionic association is entropically unfavourable (since it requires highly 'distorted' electron densities to be retained for long periods of time) and will result to proton-exchange reactions; however they may constitute important intermediates, e.g. in gas-phase hydrogen-deuterium exchange reactions[39]. Self-solvation (formation of strong **intramolecular** interactions by polar and charged moieties) becomes a major driving force for gas-phase polymer folding upon removal of interactions with polar solvent molecules; likewise, hydrophobic effects vanish in the absence of water. Such considerations support a reversed order of stability for intramolecular interactions of vapourised and solvated peptides; this hypothesis is consistent with both simulations presented throughout this thesis and is supported by findings such as the recently observed dissociation patterns of haem-bound Cytochrome C[40,41]. Such behaviour clearly distinguishes unsolvated systems from solvated ones.

Is the gas phase a biologically relevant environment? Breathtaking coulombic interactions, self-solvation and no hydrophobic effects – how can the gas-phase conformations of biomolecules be biologically relevant? Even if we choose to ignore the increasing availability of gas-phase techniques, it quickly becomes obvious that the gas phase is very similar to a "hydrophobic" environment such as that which proteins and other biomolecules encounter in membrane bilayers. The structures of membrane-interacting biopolymers have long been shrouded in mystery due to their

difficulty to crystallise and their structural flexibility. Tracing such systems in small concentrations in a polar solvent may give a misleading picture; from this perspective gas-phase structures are very promising. Most biochemical research relies on experiments that are carried out in dilute aqueous solution, an environment significantly different from the emulsion-like protoplasm (which is characterised by a dielectric constant far lower than 78) in which biopolymers function *in vivo*. The vapour phase can be conceived as the "other extreme" in the array of experimental conditions, which already consist of dilute solutions and the solid phase (such as the crystalline environment). Biochemical sciences are fundamentally *reductionist;* gas-phase research does anything but contradict this trend.

1.5.4 Elementary Concepts of IMS and Experimental Determination of Orientationally-averaged Ionic Collision Cross Sections.

Ion mobility experiments of interest are performed on mass spectrometers supplied with a collision cell which is filled with several Torr of a neutral buffer gas (usually helium). A uniform and relatively weak electric field **E** is applied throughout the cell of length ℓ (**V** is the voltage):

$$\mathbf{E} = \frac{\mathbf{V}}{\ell} \qquad \{1.21\}$$

Whilst in the drift tube, ions experience two forces: a force directly proportional to the electric field [$\mathbf{F} = q\mathbf{E}$, where q is the charge of the ion] which accelerates the ions between collisions with neutrals, and a force resulting from the collision itself. Overall, this latter force will have a decelerating effect, causing the ions to drift with an apparent constant velocity $\mathbf{v}_D$, which is directly proportional to the electric field and can calculated easily given the drift time, t_D:

$$\mathbf{v}_D = \frac{\ell}{t_D} = K\mathbf{E} \qquad \{1.22\}$$

The proportionality parameter K is the ion mobility, the value that is measured experimentally. An expression for the ion mobility at low fields is given by kinetic theory[42]:

$$K = \frac{3}{16}\frac{q}{N}\left(\frac{1}{m_a}+\frac{1}{m_b}\right)^{1/2}\left(\frac{2\pi}{k_B T}\right)^{1/2}\frac{1}{\Omega_D} \qquad \{1.23\}$$

N being the buffer gas number density, q the ionic charge, m_a and m_b the masses of the analyte ion and the buffer gas respectively, and Ω_D the diffusion momentum transfer integral, or orientationally averaged cross-section, which is a good indicator of the shape of the ions, on which the relative frequency and nature of the collisions with the buffer gas particles ultimately depend. The relations between drift velocity and mobility as described in equation {1.22} is fundamental for ions undergoing elastic collisions with neutrals and a very similar expression can be derived by taking into account conservation of momentum. The topic is discussed further in Appendix B.

The experimentally measured parameter is the arrival time t_a, the time between the creation of a pulse of ions and their arrival at the detector, and includes the drift time t_D, and some "dead time", i.e. time the ions spend *outside* the drift cell, t_0:

$$t_a = t_D + t_0 \qquad \{1.24\}$$

Since t_0 is an undesirable source of error, a method is needed to derive the mobility without including it to the drift time. For this reason the reduced mobility, K_0, which is a function of temperature, pressure and "standard" temperature and pressure, is typically defined:

$$K_0 = K \frac{P}{760\text{Torr}} \frac{273.15\text{K}}{T} \qquad \{1.25\}$$

Obviously at "standard" temperature and pressure K and K_0 coincide. Combing expressions {1.22} and {1.25}, solving for t_D and substituting into {1.24} gives:

$$t_a = \frac{1}{K_0} \frac{\ell^2 273.15\text{K}}{T 760\text{Torr}} \frac{P}{\mathbf{V}} + t_0 \qquad \{1.26\}$$

Thus plotting the arrival time versus $P/\mathbf{V}$ (where $\mathbf{V}$ is the voltage) will yield a straight line with a slope inversely proportional to K_0 intersecting the time-axis at t_0, a convenient way of factoring out t_0. Thus the mobility of the ions under study can be obtained experimentally allowing the average collision cross-section to be calculated from equations {1.22-1.26}.

The method of analysis outlined above can be applied using an *average* arrival time from experimental arrival time distributions. However, when ions of identical m/z contain two or more subspecies partially resolved in the drift tube, an alternative method for obtaining ionic mobilities is called for. Fitting the arrival time

distributions (ATDs) to an expression for ion flux provide a different methodology for deriving mobilities from experimental drift-tube data.

Sharp ion pulses injected in drift tubes are subjected to two concurrent processes; *drift* along the direction of the applied filed and *diffusion* due to thermal fluctuations, allowing Fick's law to be written as

$$\mathbf{J} = \mathbf{v}_D n - D\nabla n \quad \{1.27\}$$

J representing ion flux (current), *n* ion number density and *D* the diffusion tensor. Assuming that depleting, but no accumulating, reactions may take place with rate $k\_$, ion transport can be described by this continuity equation:

$$\frac{\partial n}{\partial t} = -\nabla \mathbf{J} - k\_ n = D\nabla^2 n - \mathbf{v}_D \nabla n - k\_ n \quad \{1.28\}$$

Note that in the above expression D is conceived as being independent of ionic position; furthermore longitudal and transverse diffusion coefficients are thought of as being nearly the same. Both conditions are fulfilled when ions drift under the influence of a uniform and sufficiently weak field. At the vanishing-field limit, the mobility *K* and the diffusion coefficient *D* are related via the Nernst-Einstein-Townsend relation:

$$K = \frac{qD}{k_B T} \quad \{1.29\}$$

which is derived in the Appendix B. Moseley *et al.*[43] have solved eq. {1.28} for *n* using an initial condition of an infinitely thin disk of ions with radius same as the one for the entrance aperture r_0 of the drift tube and centre at *z* = 0 (*z* being the direction of the field). Solving {1.27} for ions along the z-axis only yielded

$$J = \frac{ae^{-\frac{(z-v_D t)^2}{4Dt} - k\_ t}}{4\sqrt{\pi D t}} \left(v_D + \frac{z}{t} \right) \left(1 - e^{-\frac{r_0^2}{4Dt}} \right) \quad \{1.30\}$$

In the above expression *a* represents the number of ions utilised for generating one ATD. Multiple species drifting independently can be modelled by {1.30} by the sum of currents resulting for ions with different drift speeds (v_D) and diffusion coefficients (*D*).

A necessary step to complete the analysis is an algorithm to evaluate the cross-sections from a given set of nuclear coordinates for the system under study.

1.5.5 Calculation of the Average Collision Cross-Section

As explained above, the mobility of an ion drifting in a chamber of neutral gas molecules will depend on the energy transferred between the neutrals and the ions during collisions, which in turn is dependent on the geometry of the ion. This quantity is present in {1.23} in the form of an orientationally averaged momentum transfer (or cross-section) integral, Ω_D: a sum over the energy transfer for all collision events, averaged over all collision geometries. As described in[44,45] this average collision cross section is equal to

$$\Omega_D^{(1,1)} = \frac{1}{4\pi^2}\int_0^{2\pi} d\theta \int_0^{\pi} \sin\phi d\phi \frac{\pi}{8}\int_0^{2\pi} d\gamma \left(\frac{\mu}{k_B T}\right)^3 \int_0^{\infty} e^{-\frac{\mu v_r^2}{2k_B T}} v_r^5 dv_r \int_0^{\infty} 2b\left(1-\cos\chi(\theta,\phi,\gamma,v_r,b)\right)db \quad \{1.31\}$$

where θ, φ, and γ are the three Euler angles that define the orientation of the collision, v_r is the relative speed of the ion and the buffer gas particle, b is the impact parameter (describing the paths of the ion and the colliding particle *before* the collision), χ is the scattering angle (which depends on all previous parameters) and μ is the reduced mass of the ion and the buffer gas: $\mu = \frac{m_a m_b}{m_a + m_b}$, and the superscript (1,1) indicates that binary ion-neutral collisions are considered, i.e. that consecutive collisions are treated as independent events. Equation {1.31} can be solved analytically for geometrically simple objects; for two colliding hard spheres it has a solution of πb_{min}^2, b_{min} being their contact distance. However, for more complex objects such as large atomic clusters and macromolecules, numerical solutions are usually preferred. These are achieved by iterative simulation of collisions on different orientations of the input structure until a converged value has been obtained. Three dominant strategies have been described for evaluating the average collision cross-section computationally:

1. The ***trajectory method***[44] (TJ) is the most accurate method and solves the entire expression {1.31} taking into account both the interaction potential between analyte species and buffer gas particles and the effect of multiple scattering (i.e. a buffer gas particle encountering the ion several times during a collision event). It is also computationally very intensive, rendering it inappropriate for examining a large

number of structures; the methods listed below are much less 'labour-intensive' and can offer good quantitative correspondence to experimental results after careful empirical calibration.

2. In the ***exact hard sphere scattering*** model[45] (EHSS) {1.31} reduces to

$$\Omega_{\mathrm{EHSS}} = \frac{1}{4\pi^2}\int_0^{2\pi} d\theta \int_0^{\pi} \sin\phi d\phi \frac{\pi}{8}\int_0^{2\pi} d\gamma \int_0^{\infty} 2b\left(1-\cos\chi(\theta,\phi,\gamma,b)\right)db \quad \{1.32\}$$

which discards the effect of the analyte-buffer gas potential energy (thus being much more efficient) but keeps track of multiple scattering events (by calculating the scattering angle χ after every simulated collision). This method performs best for large systems, where the effect of the ion-neutral interaction is not as great as multiple scattering. It has also been found that it may occasionally overestimate the effect of multiple scattering, resulting to larger values than the TJ method.

3. The ***projection approximation***[46-48] (PA) is historically the first methodology to be employed to validate molecular structures from diffusion properties. In E. Mack's[46] early implementation in 1925 shadows of wax models of molecules studied were projected onto a calibrated screen using light produced by a lantern, several projections being necessary to estimate an average projected area. Nowadays it remains a broadly adopted method, being the simplest and computationally 'cheapest' algorithm. Each candidate conformation is described as a collection of overlapping hard spheres and the collision cross section is approximated by their 2-dimensional projection, averaged over all possible geometries (in practice, sufficient different orientations to achieve convergence to a pre-specified level of accuracy). Thus the collision integral effectively becomes

$$\Omega_{\mathrm{PA}} = \frac{1}{4\pi^2}\int_0^{2\pi} d\theta \int_0^{\pi} \sin\phi d\phi \frac{\pi}{8}\int_0^{2\pi} d\gamma \pi b_{\min}^2 \quad \{1.33\}$$

This rendering does not take into account multiple scattering effects nor ion-neutral interactions. Yet the method was improved by[48] , who employ a 12-6-4 Lennard-Jones potential to account for the ion-neutral interaction effects. This modification is implemented in the *sigma* program. This method performs well for smaller systems, especially those with rounded, convex shapes; for largely concave objects

(such as cups) the method will provide smaller cross-section estimates[45] since for such systems multiple scattering becomes important.

The above methods outline how the orientationally averaged collision cross section can be estimated theoretically, thus establishing the link between theory and experiment that we are seeking.

1.5.6 Ion Mobility Mass Spectrometer: Elementary Design and Resolution Considerations

Drift tube experiments have a long and history from their creation until modern times. Ion mobility was used initially as a stand-alone device for analysis of atomic and small molecular gas-phase ions and their reactions[12,49]; yet in conjunction with MS, almost certain identification of ionic species by its m/z ratio has rendered the technique much more powerful. A review of the early usage and development of IMS techniques are beyond the scope of this study; Chapter 2 from Mason and McDaniel's "Transport properties of Ions in Gases"[12] as well as Clemmer and Jarrold's review article[49] both contain short but amply referenced overviews of the method's history until the late 1990's that can be consulted for further information on the subject.

An ion mobility mass spectrometer must at least comprise of the following parts:

- An ion source, followed by
- an accumulation/transfer region (not necessary if ions are produced directly in the drift region) guiding and pulsing ions into the
- drift tube, in which mobility separation occurs; ions subsequently move via the
- exit optics to a
- mass analyser, after which ions move through
- transfer optics to reach an
- ion detector.

The aforementioned IMMS architecture is minimal and different permutations even in this basic architecture are possible, e.g. positioning a mass filter before the drift cell, instead of after. The modularity of mass spectrometers in general have allowed for a range of more complex designs to be realised (see[50,51] and references therein),

some of which will be referred to later in this section. Ions, once produced in the ion source must be transferred into the high-pressure drift region. If regions of lower pressure intervene between these two parts, ions must possess sufficient energy to overcome the outward flow of gas from the drift cell's entrance aperture. A potential difference between the accumulation region and the drift region must be imposed for such a scenario to be realised, which is termed injection energy. As it is practically difficult to balance the deceleration of ions due to the pressure differential with the acceleration due to injection energy, ions are likely to enter the drift cell with somewhat elevated kinetic energy. However, thermalisation is very rapid once the ions have entered enter the low-field drift space[12,52]. Injection energy can also be utilised as a tool for structural investigation of ions. By monitoring the injection energy ions can be collisionally activated by controlled amounts as they enter the drift cell and resultant changes in mobility can be measured directly afterwards. Numerous examples using this methodology have been reported for protein systems[53-60], a subset of which will be reviewed in the following section.

The resolving power of ion mobility apparatuses affects directly their applicability. Using a solution to the transport equation (expression {1.30}) and an expression for the ionic mobility (e.g. equation {1.23}; see also section B of the appendix) it can be shown[61,62] that the resolution $t/\Delta t$ can be approximated by

$$\frac{t}{\Delta t} \approx \left(\frac{qE\ell}{16 k_B T \ln 2} \right)^{1/2} \qquad \{1.34\}$$

IMS resolution can be improved by decreasing the temperature (a feature of some temperature-controlled drift tubes), increasing the drift field or increasing the length of the drift tube. The applied field, *E*, is limited by the density of the buffer gas, *N*, as the maximum field that can be applied must satisfy the low-field condition; to increase *E* further, *N* must be likewise increased. These considerations have lead to the development of two varieties of linear-field IMS devices coupled to mass spectrometers. **Low resolution drift cells** operate at low pressures and necessarily proportionally low fields – they are more versatile and can be integrated with greater ease into novel arrangements or existing mass spectrometers. The device used for experiments described in the following chapters is one such low-resolution device. **High-resolution drift cells**[63] operate at pressures similar to atmospheric pressure

and employ drift potentials of the kV scale. These devices cannot be inserted in a low-pressure chamber due to the high pressure differential that would otherwise result. Therefore, the source must be intimately coupled to the drift region, an arrangement that also complicates buffer gas purity. Increasing the length of the drift region is an attractive approach for improving peak separation. However the majority of the ion swarm becomes too dilute to be detected at achievable experiment times, as most of the ions discharge at the walls of the cell as radial spread surpasses by much the radius of the exit aperture in the absence of ion focusing lenses inside the drift region. Nevertheless, with the pairing of such ion-focusing methods, such as ion funnels[64-67] or an RF-field applied simultaneously to the drift field[51] this approach came within reach. This advancement not only generated high-resolution designs which used lower pressures, but also led to the development of long 'tandem'-IMS tubes[68-71] (with different sections of the tube being separated by an ion funnel) and circular drift-tube topologies[51] allowing theoretically infinite separation and separation-activation cycles.

Despite its long 'life', technological improvements, such as these described, continue to take place, driven by the need to analyse complex systems with greater accuracy. Even this very partial account of instrumental developments in the field of ion chromatography / ion mobility spectrometry would suffer if tribute is not paid to relatively novel techniques utilising non-uniform fields[72-77]. Non-uniform fields have not only been employed as an extension to existing low-field IMS methods[74], but also utilise ion transport properties at high fields for separation. Two notable cases are field-asymmetric-waveform IMS (FAIMS) and travelling-wave IMS (TWIMS).

During FAIMS[72,73,78-81] ions move via a gaseous medium between two plates on which an asymmetric alternating potential is applied. The field waveform is usually chosen so that the integral over a complete cycle vanishes; however this is achieved by applying a weak potential over a certain time interval and then reversing the direction of the field but also increasing its magnitude and proportionately decrease the time it is applied, before the field direction is reversed again and the cycle recommences. FAIMS exploits the non-linear dependence of drift speed on acceleration field which prominent at high fields for ion separation. The asymmetric waveform will result in most ions discharging at the electrodes. Only slow ions for

which the mobility is field-invariant at the applied waveform amplitude will be transmitted through the FAIMS separator[72,73] whereas other species, especially those whose mobility is strongly field-dependent, will tend to move towards the high-field plate. By scanning over a range of waveform amplitudes separation between ions with different high-field mobilities becomes possible. FAIMS adds to the power of gas-phase separation science and can make a major contribution in the analysis of complex mixtures[82-86] by adding an extra dimension that is easily controlled. Instrumental developments have thrived is this area, with FAIMS devices being integrated with IMS-mass spectrometers[83] and an ESI-FAIMS-MS employing a Paul trap mass analyser has become commercially available from Thermo-Finnigan. However high-energy collisions experienced during FAIMS have been shown to distort the structures of gas-phase peptide ions[87,88]; therefore this technique can be used for meaningful studies of molecular conformation only if detailed models of ion motion in this setup are taken into account. TWIG ion mobility spectrometry (TWIMS), developed in R. Bateman's group[76,77,89], has been widely used recently due to its commercialisation, despite the fact that a mathematical description of its operation did not appear in the literature until recently[90]. The TWIMS separation device is superficially similar to that of low and uniform-field IMS in that it comprises of a set of stacked ring electrodes. However, the applied field does not vary uniformly in space and also varies in time. In contrast to conventional IMS, the voltage 'scans' the electrodes sequentially in 'waves', of predetermined amplitude and frequency, between which no field is applied; ions can be thought to be 'surfing' on these waves. Assuming ions have been injected into the drift region, with no drift applied, their motion will be entirely thermal. As the wave approaches the instantaneous force applied to the ions can be given by the spatial derivative of the wave. "Sharp" waves with high amplitude will result to a high ionic velocity at the direction of wave propagation. If the ionic velocity exceeds the wave speed, a single wave will transfer the ions to the end of the drift tube and no separation will occur. Yet if the opposite is true (meaning broader waves with lower amplitudes propagating at a higher speed) ions will initially move forward but eventually 'tumble' at the other side of the wave and will be *decelerated* in the direction of field propagation. Under such conditions ions can be separated according to drift velocity

and mobility. The main problem arising with this setup is, once again, that of field energy; during their peak velocity, collisions in the direction of wave propagation may induce the temperature of ions to increase above that of the buffer gas. Commercial distributions of this method have also included an additional RF-field to limit radial spread and thus increase ion transmission efficiency; however the effect of the RF field on ion transport and ionic internal energy and, consequently, biomolecular structure in a relatively high-pressure region has not been assessed quantitatively to date to the best of our knowledge.

IMS is a well-established technique and despite its long tradition, novel architectures continue being designed; these features imply the applicability of such analytical techniques in the study of biomolecules.

1.5.7 Some applications of IMMS in Investigating (Bio)molecular Structure

IMMS is an attractive platform for studying biomolecular conformations for several reasons. The ability to separate conformations with different average cross-sections is an obvious desirable feature, one that is not often encountered in measurements from bulk solution. The capacity to collisionally activate ions either in the drift-tube or in an additional collision cell facilitates 'on-line' structural analysis of ions. Temperature-controlled drift cells also provide an accurate and powerful tool for studying thermodynamics. Yet prior to its application to biomolecules, IMMS was already established as an analytical technique. Early IMMS experiments focused on isolated metal ions in gases[12,49]; knowledge of their transport properties provides an experimental checkpoint for the determination of ion-neutral interaction potentials. Subsequent studies showed the capability of IMS to separate electronic states of transition metal ions[52,91-94]. IMMS proved to be a truly valuable tool for physicochemical analysis, e.g. by offering insights in the formation mechanisms of atomic clusters. To synoptically illustrate this point, let it be said drift-tube experiments could separate between carbon chains, monocyclic and polycyclic rings which at correct energies form fullerenes as well as classify the shapes of assemblies (dimers, trimers) of fullerenes and fullerene fragments (reviewed by Clemmer & Jarrold[49] and Wyttenbach & Bowers[50]). Coupling of ion mobility mass spectrometers with "gentler" ionisation methods such as ESI[95,96] or MALDI[97] enabled the study of

biomolecular structures. Studies on peptides, such as bradykinin[98], and proteins such as cytochrome C[53], ubiquitin[56], apomyoglobin[55] BPTI[54] and lysozyme[99] began and are continuing to delineate a view on the behaviour of polypeptides in the absence of solvent effects. The most salient points of this work will be summarised. IMMS of multiply charged electrosprayed protein ions revealed a general increase of collision cross-section with charge, caused by protein unfolding due to repulsive coulombic forces[49,100,101]. Furthermore, low charge states were often found to be more compact than available X-ray or NMR structures[53-55,99], indicating a structural collapse taking place in desolvated protein ions. Research on apomyoglobin ions[55] showed a different unfolding rate with charge state depending on whether the ions were positively or negatively charged: anions unfolded more uniformly but also at a slower rate whereas cations exhibited a much 'sharper' transition between 5+ and 7+ charge states. Thus, intermediate charge states differed significantly in cross section, yet high charge states seemed to converge to a similar extended structure. Thus charge *location* is implicated in gas-phase coulombic protein unfolding mechanisms (further evidence for this statement is afforded from simulation studies presented in Chapters 3 & 4). Protein IMMS detected differences resulting from solution conditions and separated several conformations of electrosprayed ions. Interestingly, ions corresponding to intermediate charge states (6-15+ for cytochrome C[53], 6-10+ for ubiquitin[56,58,102], 6-11+ for lysozyme[99]) separate into different sub-populations during IMMS experiments, revealing the presence of multiple unfolding pathways, which can be thought of as either folding intermediates or configurations resulting from destabilisation of the folded state *in vacuo*. Efforts to gain greater insight into the mechanisms of biomolecular trans-conformation in a solvent-free environment have followed, and in some cases guided, technological developments in IMMS instrumentation. IMS studies on unfolding and refolding of protein ions *in vacuo* measured ATDs and estimated collision cross sections for systems at elevated energies. Activation of ions in ion mobility mass spectrometers is usually achieved in two ways: either by collisional activation by ion acceleration in a low vacuum region of the instrument, or by performing experiments at elevated buffer gas temperatures. The former approach can be carried out in a number of different ways, some of the main ones being:

- Injection energy monitoring, where the potential pushing ions into the drift region is varied. This approach cannot be employed if the ion source is directly linked to the drift tube.
- Activation inside the drift tube. For this functionality to exist, groups of ion guides within the drift tube must be electrically isolated. Ion funnels present in tandem IMS instruments [source-$(IMS)_n$-MS] may fulfil this criterion.
- Activation in a different region of the instrument. Given the modular nature of mass spectrometers, the possible design variants for this setup are practically limitless, and may include activation in the source[102], in a pre-IMS ion trap[103-105], pre-drift tube FAIMS device[83] or an ion funnel[68].

IMS utilising weak uniform fields has the advantage of being a near-equilibrium technique; therefore ion temperature can be modified in variant-temperature drift-tubes. Even at intermediate field strengths ion temperature can be estimated when thermalisation due to the drift field is corrected for[12,61]. When activation of analytes is achieved as ions enter, or are traversing, the drift tube must necessarily take place during time-scales that are necessarily smaller than drift time; if effected by elevating the temperature of the buffer gas, an elevated ion temperature during the entire drift will result; if ions are activated in a separate part of the instrument, time-scale and energy of this event can be varied. In any case, valuable information about protein structural stability in the absence of solvent effects is reaped from such studies. Injection energy studies of protein ions succeeded in effecting conformational changes upon activation. Often, such changes include elongation of the structure of intermediate charge states[49-51,101]; yet this is not always the case. For example, collision cross sections for apomyoglobin 6+ decrease with increasing injection energy[55]. Such results indicate that it is the ('sequence-specific') gas-phase PES of the ionic polypeptide that determines the structures resulting from these experiments and a structural collapse need not be excluded from the set of possible changes. Furthermore, results such as the one just mentioned support the theory that gas-phase protein ions have some 'memory' of the solution fold, a theory also supported from measurements taken on an ion mobility mass spectrometer containing a pre-drift tube ion trap in which ions can be stored for variable amounts of time. Badman *et al.*[103,105] have reported that cytochrome C ions stored in the trap for times up to 100ms

became more elongated as a function of trapping time; these structural transitions also occurred more rapidly for high charge states than lower ones. Interestingly, the authors were able to classify the resulting ionic populations into five distinct families based on drift time, the kinetics of unidirectional conversion varied with charge. Myung *et al.*[104] have detected similar qualitative trends for ubiquitin ions stored in an ion trap for multi-millisecond timescales before IMS separation. Such studies suggest that electrosprayed protein ions can adopt a metastable state in the gas-phase. More recent studies utilising novel high-resolution tandem IMMS instrumentation[59,60,106] also show that the gas phase is an excellent environment for isolating and studying polypeptide conformations. In such high-resolution instruments, IMS spectra of protein ions contain several peaks whose width is much broader than that expected from theory (equations {1.29} and {1.30}). However isolation of narrow pulses from originally broad features by means of the ion funnel 'gates' display broadening due to diffusion in agreement to the above theoretical predictions. Such data hint to the fact that solvent-free PES of polypeptides adopts a more 'glass-like' topology, possibly due to high barriers resulting from the presence of several like charges.

Applied to polypeptides, temperature controlled drift tubes also allow to determine experimentally the thermal stability of ionised peptides in addition to providing other useful insights, such as the thermodynamics of self-association and hydration[50]. Elevated drift tube temperatures have been shown to mediate unfolding in particular protein ions[54,55,57,99]. Yet this effect is not universal over accessible drift cell temperatures. An intriguing example, bradykinin (BK, sequence: RPPGFSPFR)[*], displays noteworthy temperature resistance to elongation up to 600K[98]. ATDs resulting from low-resolution IMS separation are broader than expected from the transport equation {1.30}[55,59,98,106], an effect which may result from multiple conformations with very similar cross sections, especially if they are interconverting at timescales similar to that of drift. Notably, Wyttenbach *et al.*[98] have reported that cross sections of candidate structures of BK $[M+H]^+$ ions generated from simulated annealing do not depend significantly on charge assignment (from the three isomers tested: $R1^+$, $R9^+$, and $R1^+$ $R9^+$ C-terminus$^-$). BK contains three conformationally 'rigid' proline residues, decreasing the overall number of energetically favourable

[*] See Appendix A for a list of formalisms on amino acid nomenclature

geometries. Irrespectively of whether R1 or R9 are protonated or neutral, these side-chains form the structural core of the peptide's non-covalent architecture, interacting with carbonyl oxygen atoms; the rigidity of the guanidino group implies that there exists only a limited number of geometries in which it is adequately coordinated. These features of the amino-acid composition of BK explain why different low-energy configurations are characterised by similar cross sections. Nevertheless, more recent experiments utilising 'low-pressure' high-resolution IMS on electrosprayed BK ions[107] have succeeded to separate conformations for 1+, 2+ and 3+ ions (as well as oligomeric species) from ATDs that simply appeared broadened when measured using low-resolution instrumentation.

From this fragmentary review of research on peptide IMMS which is relevant to the work presented in following chapters we may conclude that IMMS provides technique allowing detailed analysis of isolated peptide conformations at thermal energies, with the possibility of activating ions by one of the methods outlined above. Variable temperature drift tubes enable careful monitoring of peptide temperature-dependent conformational changes as well as the deduction of thermodynamic data for ion processes that affect the transport properties of analyte ions. Yet, IMS occupies a popular but still limited niche in the entire field of gas-phase methodologies that provide great insights in molecular structure. A small number of techniques, ones that have contributed significantly in understanding the conformations of desolvated peptide ions, are summarised in the following section.

1.5.8 Experimental Approaches, other than IMMS, for Understanding Gas-phase Biomolecular Structure

Numerous techniques have been developed recently that can provide information on the structure of solvent-free or partially solvated biomolecules. As with IMS, the vast majority of these experimental setups are linked to a mass spectrometer, taking advantage of the accurate mass separation of ionised biopolymers. Results from a subset of the techniques synoptically presented here are discussed in Chapter 3.

Appearance of the mass spectrum: charge-state distribution analysis

A mass spectrum of a multiply-charged (cationic or anionic) macromolecule already contains conformational information. Since – as mentioned above – coulombic repulsion leads to charge-mediated unfolding of multiply-charged globular polymers, if the (electrosprayed) analyte ion exists in two conformational states *in vacuo*, it will become observable as two distinct Gaussian charge distributions; the low-charge (high m/z) distribution corresponding to the compact conformations and the high-charge distribution to the extended ones. This phenomenon is well-documented and analysis techniques have been suggested to quantify it[29].

Hydrogen-Deuterium exchange (HDX) is a technique used frequently in conjunction with NMR spectroscopy in order to probe conformational flexibility in solution. The process causes a mass shift of 1 neutron mass for each hydrogen that is exchanged, which becomes directly observable by mass spectrometry. It also leads to broadening of isotopic distributions which can also be manipulated to obtain conformational information[108]. HDX can be used to monitor the condensed-phase structures of biomolecules if it is performed when the analyte is still in solution; but it can also be carried out *in vacuo* if the analyte ions are 'trapped' and 'stored' in the instrument in the presence of vapourised neutral deuterated solvent molecules, enabling the study of gas-phase acid-base reactions[109-112]. Several mechanisms for gas-phase HDX have been suggested and it seems possible that the precise mechanism is influenced by the structure and basicity of the deuterating agent[113,114]. The 'relay mechanism' is thought to be the most prominent pathway for gas-phase HDX involving reactions deuteron carriers with relatively low basicity such as D_2O and MeOD[114], with the exception of exchange of carboxylic acids which may proceed via the related 'flip-flop' mechanism; schematics of these mechanisms are shown in Figure 1.4.

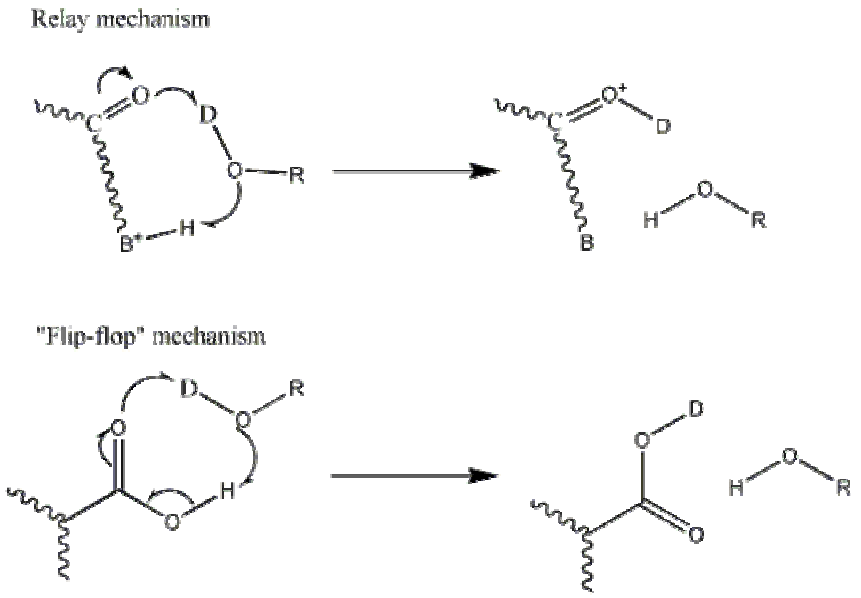


Figure 1.4
Schematics of “relay” and “flip-flop” HDX mechanisms.

During the relay mechanism the deuterating agent must be simultaneously solvated by an electron donor (specifically interacting with a D atom) and proton (interacting with the electron-donating atom of the deuterating agent, i.e. oxygen in the case of MeOD and D_2O) elsewhere in the peptide; a proton is then transferred to the deuterating agent while a deuterium atom is transferred to the peptide. In the flip-flop mechanism a carboxylic acid group must hydrogen-bond simultaneously with a deuteron and the oxygen (or other electron-donating atom) of the deuteron carrier; the deuteron is subsequently relayed to the peptide as the proton is absorbed by the deuterating agent. As charged sites excluding basic amino-acid side-chains (i.e. R, K and H) are usually characterised by a lower proton affinity, the relay mechanism is thought to be charge mediated. Hence, protonated peptides with positive charges that are equal to or fewer than the number of arginines in the sequence are expected to exhibit slower exchange rates. Yet other features of the mechanism may have a bearing on exchange rates. As proton and deuteron transfer occur almost simultaneously, exchange by the relay mechanism cannot take place unless the separation between the proton donating and accepting groups on the peptide is less than *ca.* 5Å. Furthermore, intramolecular hydrogen bonding will decrease the HDX rate. However, a low exchange rate can be observed in polypeptides which adopt highly extended configurations, e.g. due to coulombic repulsion[115,116]. Thus broad

structural characterisation can be made on the basis of gas-phase HDX experiments, however caution must be taken as both very compact and very extended structures can be characterised by low exchange rates. HDX has been successfully used to separate conformers, which appear as distinct features in isotopic distributions[113,115-117]. It must be noted though that coupling of IMS with HDX reveals a complex relationship between the two techniques. Although different exchange rates have been observed for different IMS-separated conformers in experiments where HDX was performed in a drift tube[115,116], performing IMS and HDX sequentially has shown that the techniques are largely complementary, allowing different conformations that are otherwise unresolvable to be detected even for relatively simple systems, such as BK[117].

Dissociation techniques encompass a broad class of particularly powerful ion manipulation methods. They have become commonplace in protein sequencing, but higher-order structural information can be inferred from them as well. The basis for dissociation techniques is drawn from the ability to fragment analyte ions *in vacuo*. Fragmentation and dissociation of biomolecular ions can take place by means of thermal collisions with a chemically inert gas (collision-induced dissociation, CID), bombardment with low-energy electrons (electron-capture dissociation and electron transfer dissociation ECD/ETD) or irradiation with light of a particular wavelength (such as blackbody infrared radiation dissociation, BIRD), among others. These techniques, ECD in particular, are gentle enough to be sensitive to differing energy barriers to dissociation of non-covalent interactions, which are linked to conformational energies[118]. Cleavage of both backbone and side-chains can result from such approaches; the nomenclature of backbone fragments is shown in Figure 1.5.

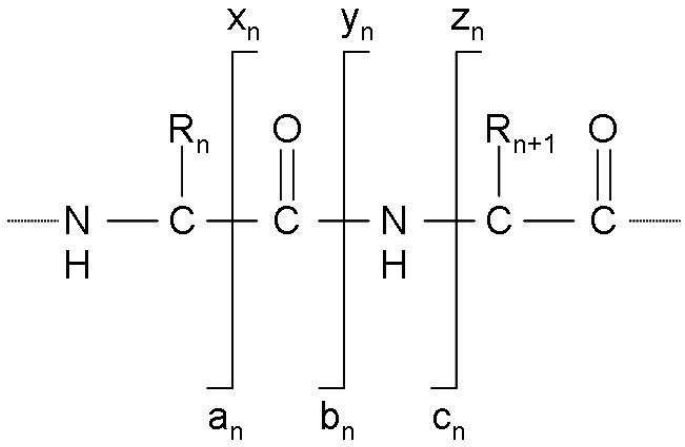


Figure 1.5

Peptide fragments resulting from backbone cleavage. Here, R corresponds to the amino acid side-chain.

In CID ions are accelerated in the presence of low pressures of neutral gas molecules until excitation is sufficient to lead to bond cleavage. Several mechanisms have been put forward for collisionally-induced fragmentation (see Paizs and Suhai[119] for a review) the majority of which are charge-mediated, even though charge-remote cleavage can also take place. Most readily witnessed fragmentation from CID is along the peptide bond, which can give rise to *b* and *y* ions. Dissociation energetics are best accounted for by the 'heterogeneous population'[120] and 'mobile proton'[121] models. Both theories implicate ions protonated at 'secondary' protonation sites, notably the backbone nitrogen atom, for fragmentation to occur. The heterogeneous population model entails the existence of a mixture of protonation states in peptide ions; the mobile proton model maintains that peptide-neutral collisions displace the proton from sites with higher to sites with lower gas-phase basicity. This is equivalent to saying that there is a temperature at which intrapeptide proton transfer reactions are frequent. The mobile proton model accounts for the correlation between dissociation energy and gas-phase basicity. Formation of *b*-*y* fragments observed experimentally is thought to be initiated by proton transfer to the peptide backbone at the site of cleavage – protonation on the backbone nitrogen, albeit thermodynamically less favourable than protonation on the backbone carbonyl group, weakens the peptide bond. Nucleophilic attack by a proximal electronegative

group to the carbon atom bound to backbone NH_2^+ to form a complex of the *b* fragment, containing a cyclic derivative involving its C-terminal carbonyl, the precise nature of which depends on the electron-donating group, and a *y* fragment. Depending on the relative proton affinities of the two peptide fragments the proton may transfer between them prior to dissociation. High-energy collisions occurring during CID will likely have an effect on peptide conformation, a hypothesis verified by the IMMS results summarised in section 1.5.7; nevertheless peptide structure beyond amino acid composition does have some effect on the fragmentation pathway[122].

Gas-phase spectroscopic techniques

Spectroscopic methods are an established tool for the study of biomolecules in the condensed phases. Therefore it is not surprising that the coupling of spectroscopic techniques such as IR[123,124] and fluorescence[125-128] to mass spectrometry has turned out to be particularly profitable for the study of small 'minimal' systems[124-127] and even much larger proteins and peptides[123,128]. Also of particular interest are two-photon ionisation / double resonance techniques which, performed on supersonic molecular jets, allow detailed analysis to be made on peptides several amino acids in length[129-137]. Spectroscopic studies have indicated conformational preferences in mono-, di- and tri-peptides and peptide analogues[129,131-135,138-146] thus pointing out with high precision the intrinsically stable structures at a local level of protein structure. Double resonance experiments have succeeded to detect gas-phase helix formation[132,147-150] and verify the greater stability of 3_{10}-helices over α-helices *in vacuo*. Overall, such efforts exhibit great interest and the development of hybrid instruments incorporating these techniques as the ones mentioned above will undoubtedly prove very fruitful in the study of biomolecular conformations.

1.6 Systems Under Study

As explained in section 1.5.3, numerous aspects of the gas-phase behaviour of biomolecules are of great interest. All work presented herein focuses on model peptide systems, with the ultimate aim of their structural characterisation. Once details of the methodology chosen have been revealed, results from the investigation of conformational preferences *in vacuo* of a well-folding model protein, the Trp cage miniprotein, will be presented. The effect of factors such as protonation state and

proton location, addition of protecting groups at the peptide termini and incorporation of amino-acid enantiomers on the gas-phase structure of isolated protein ions are studied by a variety of gas-phase techniques and *in silico* modelling. Furthermore, the thermal stability of the gas-phase structures of Trp cage ions are also traced experimentally, an investigation that is rendered possible by heating and temperature regulation of the drift tube. Studies of Trp cage focus almost entirely on the folding and conformational stability of a single peptide chain. On the other hand experiments that follow studies on Trp cage focus on the analysis of peptide assemblies. An 11-residue fragment of human protein Transthyretin (TTR) serves as a model amyloid system. Using IMMS and modelling, we perform a preliminary characterisation of oligomeric self-assemblies of this endekapeptide and discuss possible implications for the mechanism of amyloid fibril formation.

1.7 References

(1) Becker, O. M.; Karplus, M. *Journal of Chemical Physics* **1997**, *106*, 1495-1517.

(2) Wilson, C. A.; Kreychman, J.; Gerstein, M. *Journal of Molecular Biology* **2000**, *297*, 233-249.

(3) http://www.research.ibm.com/bluegene/BG_External_Presentation_January_2002.pdf 2002.

(4) Garen, A.; Levinthal, C. *Biochimica Et Biophysica Acta* **1960**, *38*, 470-483.

(5) Levintha.C *Journal De Chimie Physique Et De Physico-Chimie Biologique* **1968**, *65*, 44-&.

(6) Ponder, J. W.; Case, D. A. *Protein Simulations* **2003**, *66*, 27-+.

(7) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *Journal of the American Chemical Society* **1984**, *106*, 765-784.

(8) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *Journal of Computational Chemistry* **1986**, *7*, 230-252.

(9) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *Journal of the American Chemical Society* **1995**, *117*, 5179-5197.

(10) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P. *Journal of Computational Chemistry* **2003**, *24*, 1999-2012.

(11) Frenkel, D.; Smit, B. *Understanding molecular simulation : from algorithms to applications*; 2nd ed.; Academic Press: San Diego, 2002.

(12) Mason, E. A.; McDaniel, E. W. *Transport properties of ions in gases*; Wiley: New York, 1988.

(13) Hansmann, U. H.; Okamoto, Y. *Curr Opin Struct Biol* **1999**, *9*, 177-83.

(14) Vasquez, M.; Nemethy, G.; Scheraga, H. A. *Chemical Reviews* **1994**, *94*, 2183-2239.

(15) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671-680.

(16) Li, Z. Q.; Scheraga, H. A. *Proceedings of the National Academy of Sciences of the United States of America* **1987**, *84*, 6611-6615.

(17) Cvijovic, D.; Klinowski, J. *Science* **1995**, *267*, 664-666.

(18) Leach, A. R. *Molecular Modelling: Principles and Applications*; Pearson Education, 2001.

(19) Wales, D. J. *Energy Landscapes (With Applicacations to Clusters, Biomolecules and Glasses)*; Cambridge University Press, 2003.

(20) Becker, O. *JOURNAL OF COMPUTATIONAL CHEMISTRY* **1998**, *19*, 1255-1267.

(21) Becker, O. *PROTEINS-STRUCTURE FUNCTION AND GENETICS* **1997**, *27*, 213-226.

(22) Levy, Y.; Jortner, J.; Becker, O. *JOURNAL OF CHEMICAL PHYSICS* **2001**, *115*, 10533-10547.

(23) Mortenson, P.; Evans, D.; Wales, D. *JOURNAL OF CHEMICAL PHYSICS* **2002**, *117*, 1363-1376.

(24) Hesselbo, B.; Stinchcombe, R. B. *Physical Review Letters* **1995**, *74*, 2151-2155.

(25) Lyubartsev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; Vorontsovvelyaminov, P. N. *Journal of Chemical Physics* **1992**, *96*, 1776-1783.

(26) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Biopolymers* **2001**, *60*, 96-123.

(27) Hansmann, U. H. E.; Okamoto, Y. *Journal of Computational Chemistry* **1997**, *18*, 920-933.

(28) Sugita, Y.; Okamoto, Y. *Chemical Physics Letters* **1999**, *314*, 141-151.

(29) Dobo, A.; Kaltashov, I. *ANALYTICAL CHEMISTRY* **2001**, *73*, 4763-4773.

(30) Hudgins, R.; Woenckhaus, J.; Jarrold, M. *INTERNATIONAL JOURNAL OF MASS SPECTROMETRY* **1997**, *165*, 497-507.

(31) Shelimov, K.; Clemmer, D.; Hudgins, R.; Jarrold, M. *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY* **1997**, *119*, 2240-2248.

(32) Shelimov, K.; Jarrold, M. *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY* **1997**, *119*, 2987-2994.

(33) Valentine, S.; Counterman, A.; Clemmer, D. *JOURNAL OF THE AMERICAN SOCIETY FOR MASS SPECTROMETRY* **1997**, *8*, 954-961.

(34) Woenckhaus, J.; Mao, Y.; Jarrold, M. *JOURNAL OF PHYSICAL CHEMISTRY B* **1997**, *101*, 847-851.

(35) Schnier, P.; Price, W.; Jockusch, R.; Williams, E. *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY* **1996**, *118*, 7178-7189.

(36) Ewing, N.; Pallante, G.; Zhang, X.; Cassady, C. *JOURNAL OF MASS SPECTROMETRY* **2001**, *36*, 875-881.

(37) Price, W.; Schnier, P.; Williams, E. *ANALYTICAL CHEMISTRY* **1996**, *68*, 859-866.

(38) Wyttenbach, T.; vonHelden, G.; Bowers, M. *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY* **1996**, *118*, 8355-8364.

(39) Wyttenbach, T.; Paizs, B.; Barran, P.; Breci, L.; Liu, D.; Suhai, S.; Wysocki, V.; Bowers, M. *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY* **2003**, *125*, 13768-13775.

(40) Breuker, K.; McLafferty, F. *ANGEWANDTE CHEMIE-INTERNATIONAL EDITION* **2005**, *44*, 4911-4914.

(41) Breuker, K.; McLafferty, F. *ANGEWANDTE CHEMIE-INTERNATIONAL EDITION* **2003**, *42*, 4900-4904.

(42) Mason, E.; McDaniel, E. *Transport Properties of Ions in Gases*, 1988.

(43) Moseley, J. T.; Gatland, I. R.; Martin, D. W.; McDaniel, E. W. *Physical Review* **1969**, *178*, 234-&.

(44) Mesleh, M.; Hunter, J.; Shvartsburg, A.; Schatz, G.; Jarrold, M. *JOURNAL OF PHYSICAL CHEMISTRY* **1996**, *100*, 16082-16086.

(45) Shvartsburg, A.; Jarrold, M. *CHEMICAL PHYSICS LETTERS* **1996**, *261*, 86-91.

(46) Mack, E. *Journal of the American Chemical Society* **1925**, *47*, 2468-2482.

(47) VONHELDEN, G.; HSU, M.; GOTTS, N.; BOWERS, M. *JOURNAL OF PHYSICAL CHEMISTRY* **1993**, *97*, 8182-8192.

(48) Wyttenbach, T.; vonHelden, G.; Batka, J.; Carlat, D.; Bowers, M. *JOURNAL OF THE AMERICAN SOCIETY FOR MASS SPECTROMETRY* **1997**, *8*, 275-282.

(49) Clemmer, D. E.; Jarrold, M. F. *Journal of Mass Spectrometry* **1997**, *32*, 577-592.

(50) Wyttenbach, T.; Bowers, M. T. In *Modern Mass Spectrometry* 2003; Vol. 225, p 207-232.

(51) Bohrer, B. C.; Mererbloom, S. I.; Koeniger, S. L.; Hilderbrand, A. E.; Clemmer, D. E. *Annual Review of Analytical Chemistry* **2008**, *1*, 293-327.

(52) Kemper, P. R.; Bowers, M. T. *Journal of the American Society for Mass Spectrometry* **1990**, *1*, 197-207.

(53) Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. *Journal of the American Chemical Society* **1995**, *117*, 10141-10142.

(54) Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. *Journal of the American Chemical Society* **1997**, *119*, 2240-2248.

(55) Shelimov, K. B.; Jarrold, M. F. *Journal of the American Chemical Society* **1997**, *119*, 2987-2994.

(56) Valentine, S. J.; Counterman, A. E.; Clemmer, D. E. *Journal of the American Society for Mass Spectrometry* **1997**, *8*, 954-961.

(57) Jarrold, M. F. *Accounts of Chemical Research* **1999**, *32*, 360-367.

(58) Badman, E. R.; Hoaglund-Hyzer, C. S.; Clemmer, D. E. *Journal of the American Society for Mass Spectrometry* **2002**, *13*, 719-723.

(59) Koeniger, S. L.; Merenbloom, S. I.; Sevugarajan, S.; Clemmer, D. E. *Journal of the American Chemical Society* **2006**, *128*, 11713-11719.

(60) Koeniger, S. L.; Clemmer, D. E. *Journal of the American Society for Mass Spectrometry* **2007**, *18*, 322-331.

(61) Revercomb, H. E.; Mason, E. A. *Analytical Chemistry* **1975**, *47*, 970-983.

(62) Siems, W. F.; Wu, C.; Tarver, E. E.; Hill, H. H.; Larsen, P. R.; McMinn, D. G. *Analytical Chemistry* **1994**, *66*, 4195-4201.

(63) Dugourd, P.; Hudgins, R. R.; Clemmer, D. E.; Jarrold, M. F. *Review of Scientific Instruments* **1997**, *68*, 1122-1129.

(64) Shaffer, S. A.; Tang, K. Q.; Anderson, G. A.; Prior, D. C.; Udseth, H. R.; Smith, R. D. *Rapid Communications in Mass Spectrometry* **1997**, *11*, 1813-1817.

(65) Shaffer, S. A.; Prior, D. C.; Anderson, G. A.; Udseth, H. R.; Smith, R. D. *Analytical Chemistry* **1998**, *70*, 4111-4119.

(66) Shaffer, S. A.; Tolmachev, A.; Prior, D. C.; Anderson, G. A.; Udseth, H. R.; Smith, R. D. *Analytical Chemistry* **1999**, *71*, 2957-2964.

(67) Kim, T.; Udseth, H. R.; Smith, R. D. *Analytical Chemistry* **2000**, *72*, 5014-5019.

(68) Koeniger, S. L.; Merenbloom, S. I.; Valentine, S. J.; Jarrold, M. F.; Udseth, H. R.; Smith, R. D.; Clemmer, D. E. *Anal Chem* **2006**, *78*, 4161-74.

(69) Merenbloom, S. I.; Koeniger, S. L.; Valentine, S. J.; Plasencia, M. D.; Clemmer, D. E. *Analytical Chemistry* **2006**, *78*, 2802-2809.

(70) Merenbloom, S. I.; Koeniger, S. L.; Bohrer, B. C.; Valentine, S. J.; Clemmer, D. E. *Analytical Chemistry* **2008**, *80*, 1918-1927.

(71) Valentine, S. J.; Kurulugama, R. T.; Bohrer, B. C.; Merenbloom, S. I.; Sowell, R. A.; Mechref, Y.; Clemmer, D. E. *International Journal of Mass Spectrometry* **2009**, *283*, 149-160.

(72) Buryakov, I. A.; Krylov, E. V.; Makas, A. L.; Nazarov, E. G.; Pervukhin, V. V.; Rasulev, U. K. *Pisma V Zhurnal Tekhnicheskoi Fiziki* **1991**, *17*, 60-65.

(73) Buryakov, I. A.; Krylov, E. V.; Nazarov, E. G.; Rasulev, U. K. *International Journal of Mass Spectrometry and Ion Processes* **1993**, *128*, 143-148.

(74) Loboda, A. *Journal of the American Society for Mass Spectrometry* **2006**, *17*, 691-699.

(75) Loboda, A. V.; Kozlovski, V. I.; Chardakova, E. V.; Tolmachev, A. V.; Sulimenkov, I. V.; Dodonov, A. F.; Wollnik, H. *Rapid Communications in Mass Spectrometry* **1998**, *12*, 45-49.

(76) Wildgoose, J.; McKenna, T.; Hughes, C.; Giles, K.; Pringle, S.; Campuzano, I.; Langridge, J.; Bateman, R. H. *Molecular & Cellular Proteomics* **2006**, *5*, 78.

(77) Pringle, S. D.; Giles, K.; Wildgoose, J. L.; Williams, J. P.; Slade, S. E.; Thalassinos, K.; Bateman, R. H.; Bowers, M. T.; Scrivens, J. H. *International Journal of Mass Spectrometry* **2007**, *261*, 1-12.

(78) Viehland, L. A.; Guevremont, R.; Purves, R. W.; Barnett, D. A. *International Journal of Mass Spectrometry* **2000**, *197*, 123-130.

(79) Viehland, L. A.; Mason, E. A. *Annals of Physics* **1975**, *91*, 499-533.

(80) Viehland, L. A.; Mason, E. A. *Annals of Physics* **1978**, *110*, 287-328.

(81) Spangler, G. E.; Miller, R. A. *International Journal of Mass Spectrometry* **2002**, *214*, 95-104.

(82) Nikolaev, E. N.; Shukla, A.; Masselon, C.; Sharma, S.; Pasa-Tolic, L.; Smith, R. *Molecular & Cellular Proteomics* **2004**, *3*, S136-S136.

(83) Tang, K. Q.; Li, F. M.; Shvartsburg, A. A.; Strittmatter, E. F.; Smith, R. D. *Analytical Chemistry* **2005**, *77*, 6381-6388.

(84) Shvartsburg, A. A.; Li, F. M.; Tang, K. Q.; Smith, R. D. *Analytical Chemistry* **2006**, *78*, 3304-3315.

(85) Canterbury, J. D.; Yi, X. H.; Hoopmann, M. R.; MacCoss, M. J. *Analytical Chemistry* **2008**, *80*, 6888-6897.

(86) Saba, J.; Bonneil, E.; Pomies, C.; Eng, K.; Thibault, P. *Journal of Proteome Research* **2009**, *8*, 3355-3366.

(87) Shvartsburg, A. A.; Li, F. M.; Tang, K. Q.; Smith, R. D. *Analytical Chemistry* **2007**, *79*, 1523-1528.

(88) Robinson, E. W.; Shvartsburg, A. A.; Tang, K.; Smith, R. D. *Analytical Chemistry* **2008**, *80*, 7508-7515.

(89) Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H. *Rapid Communications in Mass Spectrometry* **2004**, *18*, 2401-2414.

(90) Shvartsburg, A. A.; Smith, R. D. *Analytical Chemistry* **2008**, *80*, 9689-9699.

(91) Kemper, P. R.; Bowers, M. T. *Journal of the American Chemical Society* **1990**, *112*, 3231-3232.

(92) Bowers, M. T.; Kemper, P. R.; Vonhelden, G.; Hsu, M. T. *Fundamentals of Gas Phase Ion Chemistry* **1991**, *347*, 55-85.

(93) Vankoppen, P. A. M.; Kemper, P. R.; Bowers, M. T. *Journal of the American Chemical Society* **1992**, *114*, 10941-10950.

(94) Kemper, P. R.; Bowers, M. T. *Journal of Physical Chemistry* **1991**, *95*, 5134-5146.
(95) Guevremont, R.; Siu, K. W. M.; Wang, J. Y.; Ding, L. Y. *Analytical Chemistry* **1997**, *69*, 3959-3965.
(96) Wyttenbach, T.; Kemper, P. R.; Bowers, M. T. *International Journal of Mass Spectrometry* **2001**, *212*, 13-23.
(97) Vonhelden, G.; Wyttenbach, T.; Bowers, M. T. *Science* **1995**, *267*, 1483-1485.
(98) Wyttenbach, T.; vonHelden, G.; Bowers, M. T. *Journal of the American Chemical Society* **1996**, *118*, 8355-8364.
(99) Valentine, S. J.; Anderson, J. G.; Ellington, A. D.; Clemmer, D. E. *Journal of Physical Chemistry B* **1997**, *101*, 3891-3900.
(100) Hoaglund-Hyzer, C. S.; Counterman, A. E.; Clemmer, D. E. *Chemical Reviews* **1999**, *99*, 3037-3079.
(101) Jarrold, M. F. *Annual Review of Physical Chemistry* **2000**, *51*, 179-207.
(102) Li, J. W.; Taraszka, J. A.; Counterman, A. E.; Clemmer, D. E. *International Journal of Mass Spectrometry* **1999**, *187*, 37-47.
(103) Badman, E. R.; Hoaglund-Hyzer, C. S.; Clemmer, D. E. *Analytical Chemistry* **2001**, *73*, 6000-6007.
(104) Myung, S.; Badman, E. R.; Lee, Y. J.; Clemmer, D. E. *Journal of Physical Chemistry A* **2002**, *106*, 9976-9982.
(105) Badman, E. R.; Myung, S.; Clemmer, D. E. *Journal of the American Society for Mass Spectrometry* **2005**, *16*, 1493-1497.
(106) Koeniger, S. L.; Merenbloom, S. I.; Clemmer, D. E. *Journal of Physical Chemistry B* **2006**, *110*, 7017-7021.
(107) Kemper, P. R.; Dupuis, N. F.; Bowers, M. T. *International Journal of Mass Spectrometry* **2009**, Article in press.
(108) Kaltashov, Y. *INTERNATIONAL JOURNAL OF MASS SPECTROMETRY* **2005**, *240*, 249-259.
(109) Campbell, S.; Rodgers, M.; Marzluff, E.; Beauchamp, J. *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY* **1995**, *117*, 12840-12854.
(110) Evans, S.; Lueck, N.; Marzluff, E. *INTERNATIONAL JOURNAL OF MASS SPECTROMETRY* **2003**, *222*, 175-187.
(111) GARD, E.; GREEN, M.; BREGAR, J.; LEBRILLA, C. *JOURNAL OF THE AMERICAN SOCIETY FOR MASS SPECTROMETRY* **1994**, *5*, 623-631.
(112) Rozman, M. *JOURNAL OF THE AMERICAN SOCIETY FOR MASS SPECTROMETRY* **2005**, *16*, 1846-1852.
(113) Campbell, S.; Rodgers, M. T.; Marzluff, E. M.; Beauchamp, J. L. *Journal of the American Chemical Society* **1994**, *116*, 9765-9766.
(114) Campbell, S.; Rodgers, M. T.; Marzluff, E. M.; Beauchamp, J. L. *Journal of the American Chemical Society* **1995**, *117*, 12840-12854.
(115) Valentine, S. J.; Clemmer, D. E. *Journal of the American Chemical Society* **1997**, *119*, 3558-3566.
(116) Valentine, S. J.; Clemmer, D. E. *Journal of the American Society for Mass Spectrometry* **2002**, *13*, 506-517.
(117) Purves, R. W.; Barnett, D. A.; Ells, B.; Guevremont, R. *Rapid Communications in Mass Spectrometry* **2001**, *15*, 1453-1456.
(118) de Hoffmann, E.; Stroobant, V. *Mass Spectrometry: Principles and Applications*; Second edition ed., 2002.
(119) Paizs, B.; Suhai, S. *Mass Spectrometry Reviews* **2005**, *24*, 508-548.
(120) Burlet, O.; Yang, C. Y.; Gaskell, S. J. *Journal of the American Society for Mass Spectrometry* **1992**, *3*, 337-344.
(121) Dongre, A. R.; Jones, J. L.; Somogyi, A.; Wysocki, V. H. *Journal of the American Chemical Society* **1996**, *118*, 8365-8374.
(122) Tsaprailis, G.; Nair, H.; Somogyi, A.; Wysocki, V. H.; Zhong, W. Q.; Futrell, J. H.; Summerfield, S. G.; Gaskell, S. J. *Journal of the American Chemical Society* **1999**, *121*, 5142-5154.
(123) Oomens, J.; Polfer, N.; Moore, D.; van der Meer, L.; Marshall, A.; Eyler, J.; Meijer, G.; von Helden, G. *PHYSICAL CHEMISTRY CHEMICAL PHYSICS* **2005**, *7*, 1345-1348.
(124) Polfer, N.; Paizs, B.; Snoek, L.; Compagnon, I.; Suhai, S.; Meijer, G.; von Helden, G.; Oomens, J. *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY* **2005**, *127*, 8571-8579.

(125) Chin, W.; Dognon, J.; Canuel, C.; Piuzzi, F.; Dimicoli, I.; Mons, M.; Compagnon, I.; von Helden, G.; Meijer, G. *JOURNAL OF CHEMICAL PHYSICS* **2005**, *122*, -.

(126) Chin, W.; Mons, M.; Dognon, J.; Mirasol, R.; Chass, G.; Dimicoli, I.; Piuzzi, F.; Butz, P.; Tardivel, B.; Compagnon, I.; von Helden, G.; Meijer, G. *JOURNAL OF PHYSICAL CHEMISTRY A* **2005**, *109*, 5281-5288.

(127) Iavarone, A.; Meinen, J.; Schulze, S.; Parks, J. *INTERNATIONAL JOURNAL OF MASS SPECTROMETRY* **2006**, *253*, 172-180.

(128) Iavarone, A.; Parks, J. *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY* **2005**, *127*, 8606-8607.

(129) Rizzo, T. R.; Park, Y. D.; Peteanu, L. A.; Levy, D. H. *Journal of Chemical Physics* **1986**, *84*, 2534-2541.

(130) Mons, M.; Lecalve, J.; Piuzzi, F.; Dimicoli, I. *Journal of Chemical Physics* **1990**, *92*, 2155-2165.

(131) Chin, W.; Dognon, J. P.; Canuel, C.; Piuzzi, F.; Dimicoli, I.; Mons, M.; Compagnon, I.; von Helden, G.; Meijer, G. *Journal of Chemical Physics* **2005**, *122*.

(132) Chin, W.; Piuzzi, F.; Dognon, J. P.; Dimicoli, L.; Tardivel, B.; Mons, M. *Journal of the American Chemical Society* **2005**, *127*, 11900-11901.

(133) Kamariotis, A.; Boyarkin, O. V.; Mercier, S. R.; Beck, R. D.; Bush, M. F.; Williams, E. R.; Rizzo, T. R. *Journal of the American Chemical Society* **2006**, *128*, 905-916.

(134) Mons, M.; Piuzzi, F.; Dimicoli, I. *Actualite Chimique* **2007**, 19-22.

(135) Stearns, J. A.; Guidi, M.; Boyarkin, O. V.; Rizzo, T. R. *J Chem Phys* **2007**, *127*, 154322.

(136) Stearns, J. A.; Guidi, M.; Boyarkin, O. V.; Rizzo, T. R. *Journal of Chemical Physics* **2007**, *127*.

(137) Guidi, M.; Lorenz, U. J.; Papadopoulos, G.; Boyarkin, O. V.; Rizzo, T. R. *Journal of Physical Chemistry A* **2009**, *113*, 797-799.

(138) Mons, M.; Dimicoli, I.; Tardivel, B.; Piuzzi, F.; Brenner, V.; Millie, P. *Journal of Physical Chemistry A* **1999**, *103*, 9958-9965.

(139) Piuzzi, F.; Dimicoli, I.; Mons, M.; Tardivel, B.; Zhao, Q. C. *Chemical Physics Letters* **2000**, *320*, 282-288.

(140) Chin, W.; Mons, M.; Dognon, J. P.; Piuzzi, F.; Tardivel, B.; Dimicoli, I. *Physical Chemistry Chemical Physics* **2004**, *6*, 2700-2709.

(141) Chin, W.; Compagnon, I.; Dognon, J. P.; Canuel, C.; Piuzzi, F.; Dimicoli, I.; von Helden, G.; Meijer, G.; Mons, M. *Journal of the American Chemical Society* **2005**, *127*, 1388-1389.

(142) Chin, W.; Dognon, J. P.; Piuzzi, F.; Dimicoli, I.; Mons, M. *Molecular Physics* **2005**, *103*, 1579-1587.

(143) Chin, W.; Dognon, J. P.; Piuzzi, F.; Tardivel, B.; Dimicoli, I.; Mons, M. *Journal of the American Chemical Society* **2005**, *127*, 707-712.

(144) Chin, W.; Piuzzi, F.; Dognon, J. P.; Dimicoli, I.; Mons, M. *Journal of Chemical Physics* **2005**, *123*.

(145) Chin, W.; Piuzzi, F.; Dimicoli, I.; Mons, M. *Physical Chemistry Chemical Physics* **2006**, *8*, 1033-1048.

(146) Gloaguen, E.; Pagliarulo, F.; Brenner, V.; Chin, W.; Piuzzi, F.; Tardivel, B.; Mons, M. *Physical Chemistry Chemical Physics* **2007**, *9*, 4491-4497.

(147) Brenner, V.; Piuzzi, F.; Dimicoli, I.; Tardivel, B.; Mons, M. *Journal of Physical Chemistry A* **2007**, *111*, 7347-7354.

(148) Stearns, J. A.; Boyarkin, O. V.; Rizzo, T. R. *Journal of the American Chemical Society* **2007**, *129*, 13820-+.

(149) Stearns, J. A.; Boyarkin, O. V.; Rizzo, T. R. *Chimia* **2008**, *62*, 240-243.

(150) Stearns, J. A.; Seaiby, C.; Boyarkin, O. V.; Rizzo, T. R. *Physical Chemistry Chemical Physics* **2009**, *11*, 125-132.

2

A Methodology for Delineating Structural Preferences of Biopolymers

2.1 Abstract

In this chapter the manner of implementation of the different techniques invoked in this research are described.

Configurations of the peptides under study were sampled by means of molecular dynamics simulations, using the AMBER software. Molecular dynamics formed the basis of all simulation techniques; a stochastic simulated annealing algorithm was also implemented in order to achieve better conformational sampling and is described to some detail herein. Numerical calculation of collision cross sections of modelled structures enabled comparisons with experiments. The instrumentation used for performing nano-electrospray mass spectrometry, ion mobility, CID, gas-phase HDX and CD, the manner in which these experiments were performed and, whenever necessary, the methods of additional data manipulation are also described.

2.2 Computational Methods

2.2.1.1 The AMBER Force Fields

All simulations using empirical potentials described here were performed using the AMBER8 and AMBER9 packages[1]; in particular the parameters included in the ff99 force field[2, 3] were most commonly chosen. In its simplest implementation, the energy of the system is calculated as a linear sum of bond stretching, angle bending, torsion and long range interaction energy terms {1.5}. By far, the most important contribution is yielded by the non-bonded interaction term, which consists of a 6-12 Lennard-Jones potential (emulating dispersion interactions) and a Coulomb potential (describing the electrostatic interactions) as demonstrated in Figure 2.1, which depicts the value of different energy contributions for 300 energy-minimised structures of a peptide:saccharide complex (net charge 3+) *in vacuo*, generated by simulated annealing (described below).

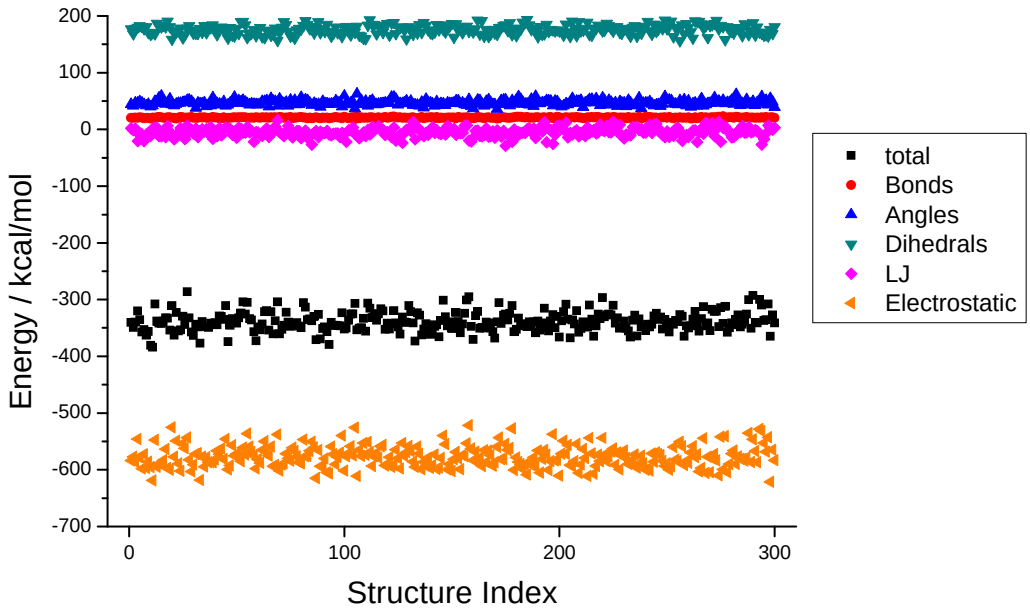


Figure 2.1 Contributions of the different energy terms to the total energy for 300 structures of the DiP1:heparin disaccharide complex (net charge = 3+) generated by simulated annealing (see section 2.2.2.4), in the absence of solvent. Evidently, electrostatic interactions contribute the most to the total energy.

2.2.1.2 Residue Development

Given the defining influence of electrostatic interactions on determining the energy of the system, care must be taken when residues that are not implemented in the AMBER package are being parameterised. N- and C-terminal amino acids with neutral termini are not included in the package thus rendering such a development necessary. In most cases described here the protocol proposed by Cornell *et al.*[2, 4] was followed (occasional exceptions will be mentioned in relevant individual sections). In brief, for each amino-acid to be parameterised, two distinct conformations were optimised at an HF/6-31G* level of theory; charges were derived from both optimised structures using the RESP fitting procedure[4]. Bond, angle and torsional parameters used for new residues existed in the force fields (since new residues commonly depicted subtle perturbations of normally occurring amino acids) so that there was no need to derive them *de novo*. The *antechamber* and *xleap* modules which are included in the AMBER package were used for the assignment of atom types (later reviewed manually) and application of the RESP fitting procedure.

2.2.2 Implementation of Conformational Sampling Strategies

As the underlying theoretical principles of the techniques used in this work, namely:

- Molecular Dynamics
- Simulated Annealing

have been outlined above (sections 1.2 and 1.3) we will now proceed with the description of our choice of the most salient options used in our simulations.

The working plan for all simulations performed can be summarised in the four following points:

I. Choice and parameterisation of initial structure
II. “Heating” / “Equilibration”
III. “Production” run
IV. Evaluation and analysis

2.2.2.1 Choice of Initial Configuration

Atomic coordinates of the starting structure where obtained 1) from a structure in the pdb (Trp cage: pdb code 1L2Y; TTR fragment 105-115 fibril model kindly supplied by Dr. Cait MacPhee but based on pdb code 1RVS) or 2) by building a high-energy extended geometry with *xleap*. All structures were energy-minimised prior to further computations, according to the protocol outlined in section 2.2.2.3. Structures in an extended configuration were submitted to several hundreds of simulated annealing iterations and the resulting minimised structures were used for performing further simulations.

In the special case of introducing one or two single-point mutations in pdb structures, which occurred for the designed Trp cage mutant peptides, the following strategy was followed.

1. Coordinates of all atoms that the two amino acids had in common were kept as in the original pdb file; the remainder were erased.
2. Missing atoms were replaced by *xleap* and optimised by the in-built minimiser
3. The peptide was parameterised and submitted to 50 cycles of simulated annealing with all atoms apart from the side-chain of the mutant amino acids being ‘locked’ at their initial position by a tight harmonic restraint.
4. The lowest-energy structure was then minimised with no restraints. The minimised structure was then used for further simulations.

All mutants treated in this way retained their backbone topology. The only exception was the Trp cage mutant containing a d-proline at position 12. In Figure 2.2 the histogram of energy microstates from a side-chain optimisation procedure is shown; notably the lowest energy structure occurs several times.

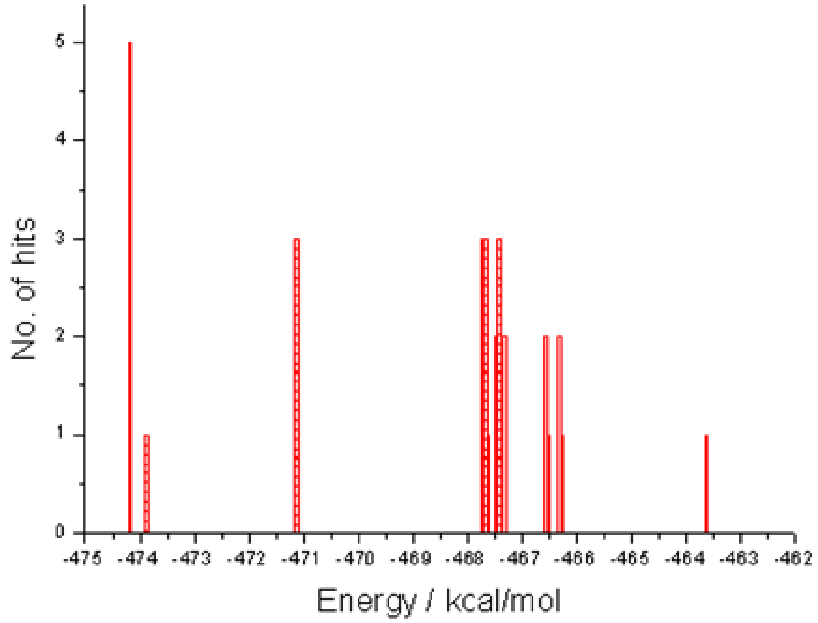


Figure 2.2 Energy histogram from the optimization of the mutated side-chain for a Trp cage construct. The lowest-energy structure is visited 25% of the time.

2.2.2.2 Molecular Dynamics Simulations

The MD simulations described herein were performed using the *sander* program included in the AMBER package. For solvent-free simulations, no cut-off was applied; in the case of an implicit solvation model the interaction cutoff distance ranged between 10 and 40 Å (the choice was usually made in conjunction with the choice of options in explicit solvent simulations, whenever that was relevant).

At the onset of an MD simulation we only have molecular coordinates in our disposition; initial velocities are assigned by the use of a random number generator but fit to a Boltzmann distribution that corresponds to the target temperature. From that point onwards, the temperature of the ensemble was regulated using the velocity Berensden thermostat[5]. Now such an assignment of atomic velocities will most likely fall in one of the higher regions of momentum space, thus a necessity to thermally "equilibrate" the ensemble arises. Using a time-coupling constant of 1 picosecond

enables thermal equilibration to occur within a few picoseconds for systems ranging from a few tens to several hundreds of atoms (Figure 2.3a, showing the absolute deviation of thermal fluctuation from the target temperature). Significant variance in the initial relaxation time can be observed and that is attributed to the initial configuration (which was not 'normalised' for the systems shown here) as well at the topography of the PES around that region. It must be noted though – and this point is illustrated in Figure 2.3b – that such rapid heating to high temperatures may have adverse effects on the initial configuration. Thus MD simulations performed to test the stability of the initial conformations were preceded by a heating scheme consisting of twelve constant temperature short runs at 25K intervals over a total time of 600ps. This protocol (used for all Trp cage MD simulations) gives rise to more gentle thermal (and conformational) fluctuations, as demonstrated by Figure 2.3c.

As described in section 1.3.1, during an MD simulation forces need to be re-evaluated in small time intervals during which the change in potential energy is essentially not taken into account. The time-step was always kept between 0.5 (for high temperature runs and all long Trp cage simulations) and 2 (for isolated low T runs) femtoseconds. A time-step of 1fs was chosen for MD during simulated annealing. To avoid bond and angle energies to contribute excessively to the energy the SHAKE algorithm[6] was enabled; bonds involving protons were kept at their equilibrium length. MD runs with excessively long time-steps would soon crash as the system would occupy unnaturally high energy levels, accompanied by excessive heating and unstable trajectories.

Solvent-free simulations represent the simplest application of the potential of equation {1.5}, the dielectric constant ε being set equal to 1. MD runs in a generalised Born solvent continuum made use of the model developed by Hawkins, Cramer and Truhlar[7] which is implemented within *sander*[8].

MD at higher temperatures involving several non-covalently bound species often required additional definitions in order to prevent the different components from drifting outside the simulation box. Such restraints normally acquired the form of flat-bottomed harmonic restraints between two appropriately chosen atoms *i* and *j*, characterised by two key distances, d_1 and d_2:

1. Whenever $r_{ij} < d_1$, $V_{restr} = 0$;
2. If $d_1 \leq r_{ij} \leq d_2$, $V_{restr} = k_2(r_{ij} - d_1)^2$, k_2 typically given a value of 5 or 10 kcal/mol
3. Finally, at $r_{ij} > d_2$, $V_{restr} = k_2[(d_2 - d_1)^2 + (r_{ij} - d_2)]$, i.e. the interaction potential increases linearly with distance.

d_1 was of the order of the length of the extended peptide and d_2 was typically below or up to 100Å. This scheme was followed for MD runs and simulated annealing of sodiated and potassiated peptides as well as for TTR105-115 multimers.

2.2.2.3 Minimisation

The parameters used during minimisation were identical to those used for the corresponding molecular dynamics runs, save the fact that SHAKE[6] was turned off and restraints, described above, were alleviated. Normally a maximum of 50000 minimisation steps was allowed per minimisation, of which the first 100 were steepest descent steps and the remainder used the conjugate gradient algorithm; the convergence criterion was $|E_i - E_{i-1}| < 0.5\,\text{cal/mol}$. All energy minimisations that were closely followed seemed to converge within these limits and in the rare occasion when this was not the case, it was attributed to a very unusual and congested initial configuration.

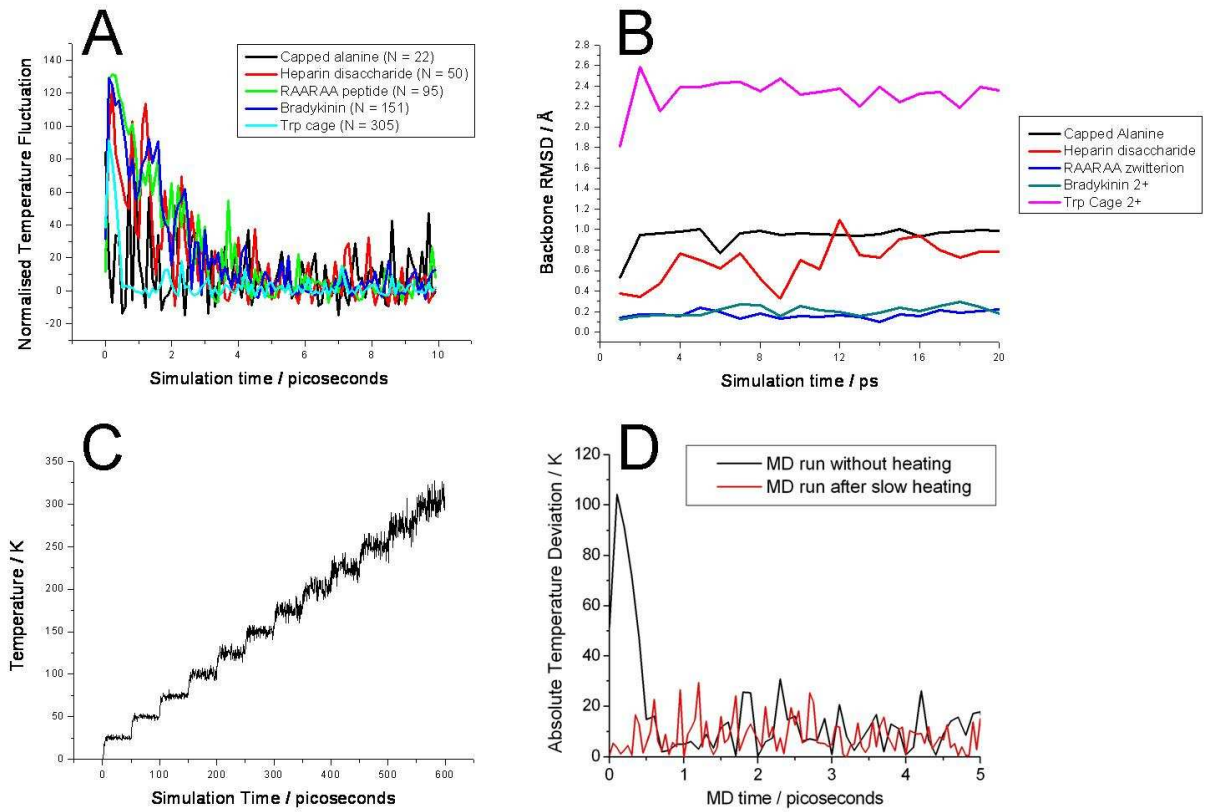


Figure 2.3 Thermal and conformational fluctuations of simulated peptides *in vacuo* upon heating. a) Temperature fluctuation for three systems each containing from 22 to 305 atoms after sudden heating to 300K, displayed for the first 10 picoseconds of the simulation.
b) Backbone RMSD (all-atom RMSD for the heparin disaccharide) for the first 20 ps of the simulation; evidently, especially for larger systems, sudden heating will perturb the initial configuration. c) Temperature as a function of simulation time, using a stepwise heating algorithm (described in the text) to heat Trp cage to 300K. d) Comparison of temperature fluctuation as a function of simulation time for a run of the Trp cage peptide undergoing sudden heating to 300K (black line) and one where the heating and relaxation algorithm shown in (c) preceded (red line).

2.2.2.4 Simulated Annealing

A simulated annealing method was used extensively for conformational sampling of the peptides under study. As reviewed in section 1.3.2 the main purpose of the algorithm[9] is to heat the system at a high temperature, at which the ergodic hypothesis is expected to hold, enabling an extensive exploration of the PES and then, by gradual decreases of temperature, essentially constraining the volume of configurations accessible to the system. The decrease in temperature is continued until the temperature is low enough so that the properties at a desirable temperature are sampled. Thus large-scale adjustments occur at higher temperatures whereas at lower temperatures more subtle adjustments take place. In our implementation of the simulated annealing algorithm MD was initially run at a high temperature (normally 800K, although for systems which were unstable at such high temperatures that was decreased to 600K) for times that ranged between 20 and 70ps, 30ps being the most common option, followed by 20 short MD runs at gradually decreasing temperature, of total simulation time that ranged between 10ps and 50ps, normally set at a value of 20 or 30ps. We consistently applied a *linear* cooling schedule; a study by Randelman and Grest[10] demonstrated for the "classical" travelling salesman optimisation problem that similar results are obtained by linear and exponential cooling schedules. When the system reached 0K the resultant peptide was energy-minimised. At 0K the free energy surface is expected to collapse to the PES and at equilibrium only the global potential energy minimum to be populated.

A drawback of this algorithm is that the system may still be trapped in an attraction basin at intermediate temperatures, especially if the global minimum changes with temperature[11] or if the PES contains several energy funnels, separated by high energy barriers. As it can be readily seen from results of re-iterated simulated annealing cycles, instead of only obtaining the expected global minimum, entire energy distributions result. Furthermore, as shown in Figure 2.4 for a model alanine-based peptide, the energy distributions depend both on the high temperature "relaxation" time and the cooling time as well. Moreover, as can be seen in Figure 2.5a for 1000 simulated annealing iterations on bradykinin the average energy fluctuates even after statistical

convergence is fulfilled, indicating the system entering different regions of the energy landscape as the run progresses. To probe the convergence characteristics of the simulated annealing algorithm, five sets of 200 simulated annealing iterations were run, each with a different cooling time (namely, 2, 10, 20, 100 and 200 picoseconds) for three different systems: the RAARAA hexapeptide (1+ zwitterion, 95 atoms), bradykinin (2+ zwitterion, 151 atoms) and the "original" Trp-cage construct (2+ charge state, 305 atoms). The results are summarised in Figure 2.5. Grest *et al.*[12], while investigating the behaviour of such surplus (residual) energies (easily conceptualised as the energy difference between the estimated energy and the minimum energy of the system: $\tilde{E} \propto \bar{E} - \min[E]$) in Ising spin-glasses, detected a reduction in such energies roughly proportional to the inverse logarithm of the cooling time, t_{cool}. In a similar vein, following a theoretical approach applicable to optimisation of many-body systems, Huse and Fisher[13] suggested

$$E(t_{cool}) \propto (\ln t_{cool})^{-\zeta} \qquad \{2.1\}$$

whereby ζ is a system-dependent variable. Figure 2.5c, displaying the residual energy in terms of (1-⟨*E*⟩/min[*E*]) as a function of cooling time (min[*E*] being the energy of the global minimum identified from all runs, so that the system converges as $\tilde{E} \to 0$), is in agreement with these observation; the lines display theoretical fits of the form

$$\tilde{E}(t_{cool}) = A/(\ln(t_{cool} + t'))^{\zeta} \qquad \{2.2\}$$

Evidently, the fits are quite acceptable; they yielded a useful set of "benchmarking" parameters for systems studied herein, the values of which are listed in Table 2.1.

Table 2.1
Convergence parameters of simulated annealing for three peptides *in vacuo*

System	*A*	*t'*	ζ
RAARAA (N=95)	3.674 ±8.986	11167 ±14934	2.033 ±1.006
Bradykinin (N=151)	5.540 ±3.675	5728.3 ±2468.0	2.086 ±0.277
Trp Cage (N=305)	205.71 ±25.32	8740.85 ±362.19	3.260 ±0.052

Parameter t' is necessary in order to represent the energy of the system at the initial state before cooling commences (that should be the energy of the system at high temperature

dynamics in terms of 1-⟨*E*⟩/min[*E*]). From these parameters we can also estimate the cooling MD time needed to satisfy particular convergence criteria. The cooling MD time needed to reach 10%, 1% and 0.5% convergence in terms of the detected global energy minimum are tabulated in Table 2.2 along with the initial energies.

Table 2.2
Initial energies and predicted simulation times required for convergence

System	$\tilde{\tilde{E}}(t_{cool}=0)$	$t_{cool}(\tilde{\tilde{E}}=0.1)$	$t_{cool}(\tilde{\tilde{E}}=0.01)$	$t_{cool}(\tilde{\tilde{E}}=0.005)$
RAARAA	0.0393	NA	85.6 ns	143 μs
Bradykinin	0.0614	NA	933 ns	1396 μs
Trp Cage	0.155	23.5 ps	1370 ns	200 μs

Several interesting features of the optimisation algorithm are revealed by such calculations. First, and somewhat unsurprisingly, the performance of simulated annealing does depend on the size of the system. Furthermore, this dependence is not linear, yet the changes in the rate of convergence are more subtle and depend on the topology of the PES of the systems studied. The asymptotic approach of the energy average to the energy of the global minimum at long simulation times is known to be due to the presence of low-lying minima separated from the ground state by small barriers[13] – the presence of several energy funnels will also impede efficient relaxation to the global minimum; however large energy barriers can be dealt with by slow cooling schedules. A worst-case scenario would be given by "glassy" PES topologies, where several minima of similar energy are separated by high barriers. The initial energies of the different systems portray the relative complexity of the PES. Many states of the relatively simple hexapeptide are expected to lie very close in energy. This is in contrast to the Trp cage miniprotein which possesses more high-energy states in which it can be trapped in at small simulation times. It is due to such differences that effects such as the very steep relaxation of Trp cage compared to the other two peptides are observed. Attention must be paid though to the fact that the *average* energy translates only indirectly to configuration; for (what is suspected to be) the global minimum occurred consistently in longer-time runs much for the RAARAA hexapeptide and bradykinin. In

general, it is necessary to repeat the procedure in order to obtain the 'fittest' structure. Clearly, the above observations indicate that simulated annealing should not be used for *global* optimisation of more complicated systems, since the simulation timescales in order to achieve "good" performance are too large and the procedure becomes inefficient. Nevertheless, simulated annealing (at more manageable simulation times) provides a sample of quasi-optimised configurations which can prove valuable as long as their validity is tested by another independent method, whether that is further simulations or experimental data. Typically each simulated annealing optimisation would involve several hundred re-iterations of the simulated annealing routine. The resultant energetics of the algorithm pose an interesting problem, given that the data generated by the optimisation are used for comparison with experimental data, namely a *rescaling* scheme must be considered in order to obtain estimates of experimental observables from the simulated ensembles. This issue will be addressed further in section 2.2.3.1, however it can be said that there is no definite solution to this problem if limited to simulated annealing instead of employing a different conformational sampling technique. Some alternatives are mentioned in Chapter 1.

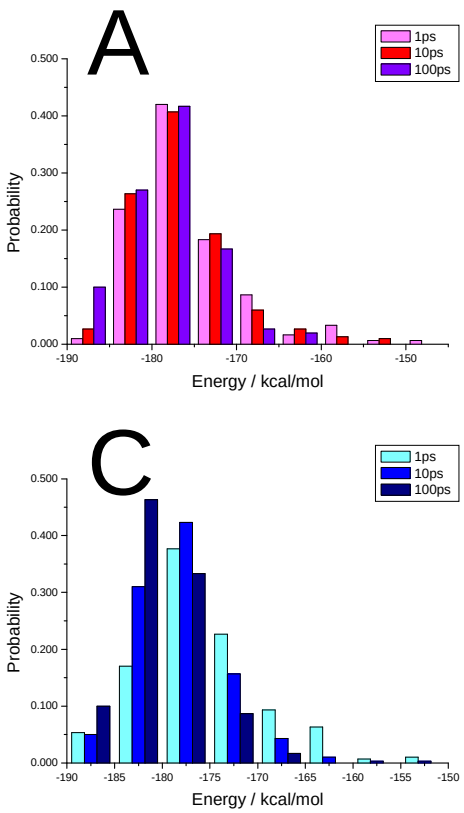


Figure 2.4 Energy distributions of 300 simulated annealing iterations for the Ace-Ala_4-NMe peptide, varying a) the length of the high-temperature MD between 1 and 10ps b) the number of cooling steps of the cooling schedule c) the length of the cooling MD between 1 and 10ps. Varying the number of cooling steps composing the cooling schedule seems to have only a minor effect on the energy distributions compared to the effect of high-T relaxing dynamics and, most importantly, the length of the cooling schedule.

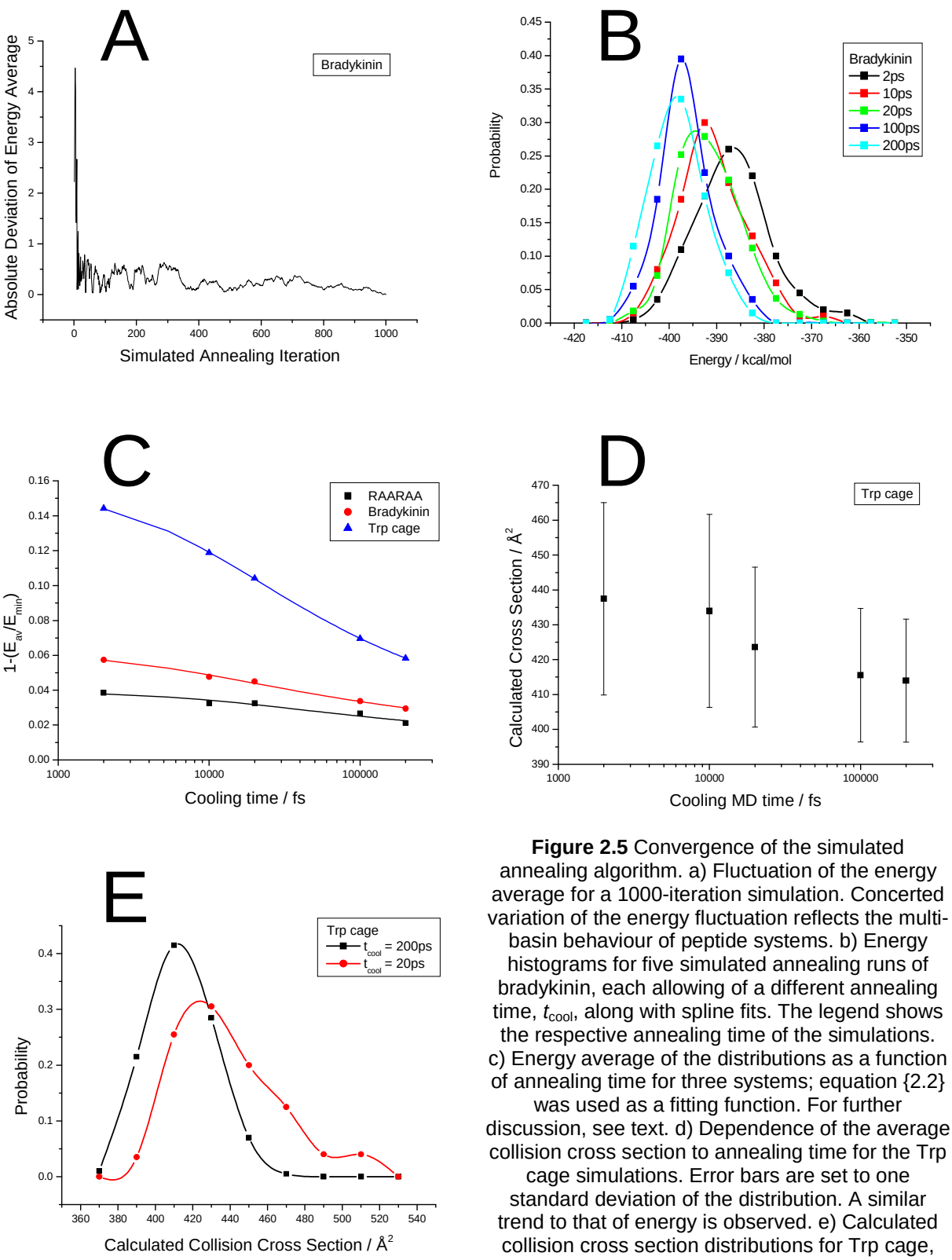


Figure 2.5 Convergence of the simulated annealing algorithm. a) Fluctuation of the energy average for a 1000-iteration simulation. Concerted variation of the energy fluctuation reflects the multi-basin behaviour of peptide systems. b) Energy histograms for five simulated annealing runs of bradykinin, each allowing of a different annealing time, t_{cool}, along with spline fits. The legend shows the respective annealing time of the simulations. c) Energy average of the distributions as a function of annealing time for three systems; equation {2.2} was used as a fitting function. For further discussion, see text. d) Dependence of the average collision cross section to annealing time for the Trp cage simulations. Error bars are set to one standard deviation of the distribution. A similar trend to that of energy is observed. e) Calculated collision cross section distributions for Trp cage, generated by simulated annealing with a cooling time of 200ps (black squares) and 2ps (red circles).

2.2.3 Performance of Different Algorithms for Estimating the Momentum Transfer Cross Section, Ω, Numerically

The three numerical integration approaches for calculating the momentum transfer integral, i.e. the collision cross-section, namely:

- The trajectory method (TJ)[14]
- The hard-sphere scattering model (EHSS)[15]
- The projection approximation (PA)[16, 17]

have been described in Section 1.5.5. All three methods are calculated by the *mobcal* program, developed by Mesleh *et al.*[14]. The integration parameters suggested by the authors were kept during the integration; these are shown in Table 2.3.

Table 2.3
Parameters for algorithms that estimate the collision cross section

Method	**TJ** *(mobcal)*	**EHSS** *(mobcal)*	**PA** *(sigma)*
Iterations	10,000	250,000	1% convergence

The performances of these settings as well as the trends followed by each algorithm are shown in Table 2.4, which displays the results from five independent *mobcal* runs for a low-energy configuration of Trp cage. As expected, PA underestimates Ω since the presence of cavities *reduces* Ω_{PA} but will tend to *augment* Ω_{EHSS} and Ω_{TJ}.

Table 2.4
Numerical Comparison and Convergence of PA, TJ and EHSS

Method	*Calculated Cross Section (Å^2)*				
PA	352.55	351.41	351.56	352.35	351.42
TJ	393.38	404.73	390.42	390.12	393.07
EHSS	407.38	406.90	405.91	406.43	404.99

Written by Gert von Helden in UCSB, *sigma*[17] is a program that calculates collision cross sections using PA and the temperature-dependent rescaling. Unlike *mobcal*, a convergence criterion is needed as input for *sigma*; this was normally set to 1% for the calculations described herein.

2.3 Mass Spectrometers

All of the experimental work reported here was performed on three commercially available instruments (two quadrupole/time-of-flight (QToF) instruments and one quadrupole ion trap), two of which have been significantly modified in-house. The original instruments were:

- A Micromass QToF-1
- A Waters QToF Ultima
- A Thermo-Finnigan LCQ

The micromass QToF-1 had been modified by addition of a high-pressure copper drift tube (followed by a transfer RF-hexapole) before the first quadrupole in order to enable ion mobility experiments[18]. The instrument in this new configuration was named Mobility Quadrupole Time-of-Flight (MoQToF). Furthermore, the source was slightly modified to allow greater control on the pressure of the first pumping stage, the region directly after the ion source which stores and transfers the ions before they are injected in the drift tube. These modifications are described in greater detail in sections 2.3.3.

The gas inlet system of the LCQ ion trap had also been altered in house. Under normal operating conditions a small amount of helium flows into the ion trap, used to perform collision-induced dissociation (CID). The helium inlet has been rebuilt to allow passage of He buffer gas from either of two vials (in addition to the original path) containing a volatile reagent which is in turn transferred into the ion trap. Gas-phase HDX can thus be performed, by filling a vial with d_4-methanol (MeOD) or D_2O (only the former was used in the studies described here). The setup will be discussed in section 2.3.4.

We shall now delve into the most salient features of the instrumentation, starting from their shared aspects and also deal briefly with the specifics of their design.

2.3.1 Nano-Electrospray Ionisation

Samples for nESI were prepared as follows. Glass capillaries (Precision Instruments Inc.; OD=1.0mm, ID=0.5mm) were pulled using a micropipette puller (Sutter Instruments Co.) and filled with the sample using gel-loading tips (Eppendorf). Alternatively (albeit seldom) pre-pulled platinum-coated glass capillaries (Promega) were used. The filled capillary was then mounted onto a capillary holder and a thin platinum wire was inserted to provide electrical contact between the sample and the power supply. Spray voltage (the potential between the capillary holder and the first electrical element of the source) was set between 1.7 and 2.8 kV for experiments on the MoQToF / QToF instruments (the most common value being that of 2.0 kV) and between 1.5 and 2.6 kV for the LCQ (a value of 1.8 was most commonly used).

Samples of near-neutral pH ("near-physiological" conditions) were buffered by 10mM ammonium acetate. All samples prepared in this way were found to have a pH between 6.2 and 6.9 (depending on the solute acidity and concentration – a pH of 6.7-6.8 was most common).

Acidified ("denatured") samples were sprayed from isochoric amounts of methanol and water, acidified with 1% v/v formic acid or 1-2% acetic acid.

Peptide concentrations ranged between 5 and 100μM for all samples apart of the aggregation TTR mixture, for which a concentration of 1 mg/ml (834μM) was used. The usage of such a high concentration of peptide was necessary to remain above the critical concentration of TTR amyloid formation[19]. TTR was prepared from water-acetonitrile mixtures containing 1% formic acid.

2.3.2 QToF Instruments

Given the similarity between Micromass (later Waters) QToF instruments, we shall describe their overall architecture in order to outline their mode of operation. The instrument can be conceptualised as consisting of the following sections:

1. Ion Source
2. Ion transfer
3. 1st mass analyser (quadrupole)
4. Collision cell
5. 2nd mass analyser (ToF)
6. Detector
7. Computerised data accumulation

A schematic of the QToF instruments is shown in Figure 2.6. We will now proceed with a more detailed description of the instrument.

QToF instruments are supplied with an off-axis "z-spray" source. nESI-produced charged droplets first encounter a sample cone positioned orthogonally with respect to the sample capillary axis. The "spray voltage" is defined as the potential difference between the sample capillary and the sample cone. An outer element is positioned outside the cone creating a gap through which heated sheath gas flows in order to assist desolvation. Upon entrance through the cone the sample begins it journey in a region of low pressure, which will gradually be decreasing to reach high vacuum. The posterior lens element is a skimmer (referred to the extraction cone or the extractor) situated again at right angle with respect to the cone, which justifies the "z-spray" label of the source topology. Skimmers effectively filter out the unfocused or "superfluous" part of the beam, owing to their shape. By positioning two orthogonally, slower (and thus, colder) ions are favoured since species with a higher momentum are less likely to succeed in completing the prescribed trajectory and remain at the centre of the particle distribution. Uncharged sample droplets are also excluded from the beam. Desolvation of the analyte begins at the tip of the Taylor cone[20] and ensues until ions reach the extractor or even later, into the first transfer element, an RF-supplied hexapole. The hexapole directly follows the extractor and, focusing the ion beam and transferring it to the analyser

chamber. The inlet region is separated from the analyser by means of a "top-hat" lens, which fulfils the double function of separating the two pumping chambers and exclude out-of-focus ions (the parts of the distribution broader than the lens' 6mm-wide aperture). In unmodified QToF instruments, the ion beam subsequently encounters a short transfer hexapole (pre-filter) which guides them to the quadrupole, the first mass analyser. Quadrupoles can be used to scan over a particular mass range but can also function as mass filters, only allowing ions at a particular m/z range to pass through[21]. Thus ions can be mass-selected at this point, a feature which is invaluable when performing CID. Ions are transferred to the collision cell by another short hexapole. The collision cell is *per se* a hexapole, yet it is enclosed into a higher pressure region, supplied by a collision gas. For CID on the QToF-1, argon (~10-15 psi, BOC gases) was used as a collision gas. The kinetic energy of the ions in the collision cell can be adjusted by varying the DC offset, RF period and amplitude. At higher energies, collisions with the neutrals will become more energetic and from a particular energy onwards peptide fragmentation will result. In MS mode, the collision cell is used just as another transfer lens, and a lower pressure, or no amount at all, of collision gas can be used. Ions are then guided into the orthogonal ToF chamber. In 'MS scan' mode they will continue their linear trajectory and hit the MS1 detector, which is a photomultiplier. In 'ToF MS scan' mode they will be repelled by a high-voltage (7.2kV) pusher pulse and drift down the ToF tube. The QToF is equipped with a reflectron, focusing the ion beam whilst diverting its path, so that the flight length is almost double that of the ToF chamber. Finally, ions are detected by the microchannel plates (MCPs). Signal is accumulated by a 4THz time-to-digital card and subsequently transferred to a computer platform where further processing occurs. Data was displayed and manipulated using *masslynx* (v. 4.0). Under this configuration the instrument contains three differential pumping regions (inlet, MS1 analyser, ToF), each pumped by a turbomolecular pump backed by a rotary pump. Modifications of the QToF-1 to obtain the MoQToF included addition of a vacuum chamber between the inlet region and the MS1 analyser.

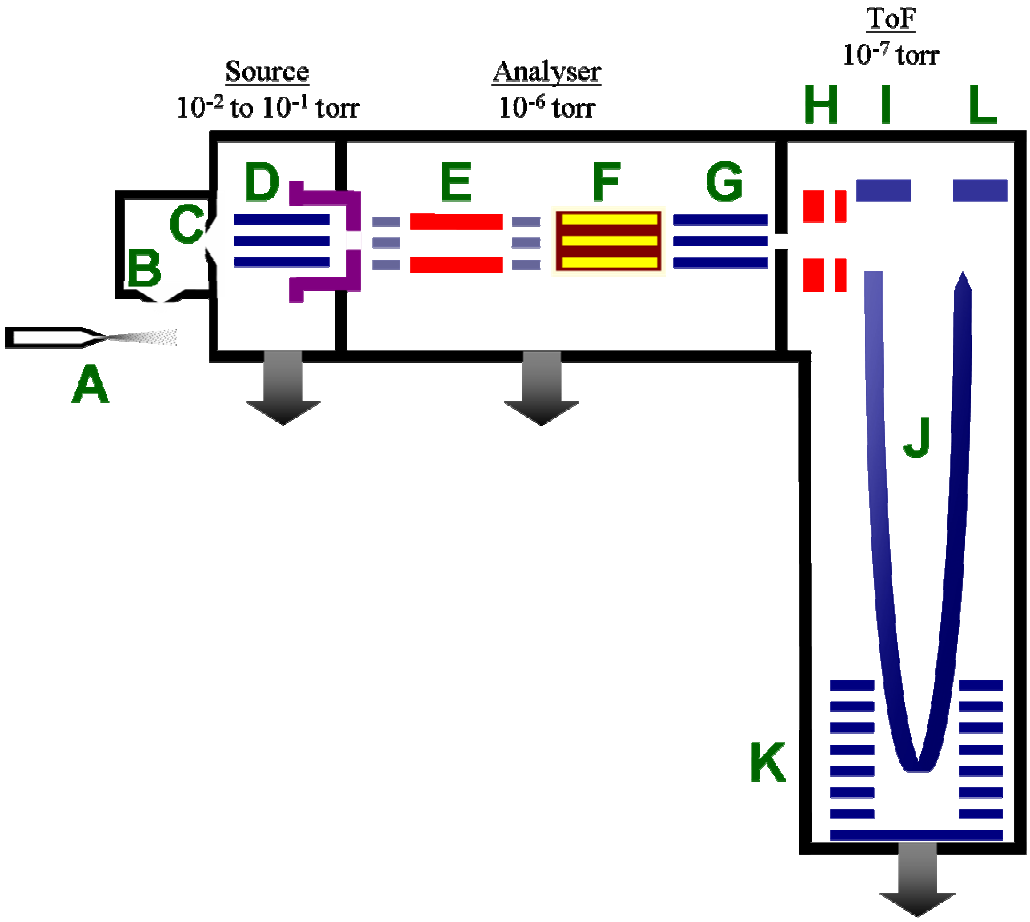


Figure 2.6 Schematic of a micromass Q-ToF. a) nESI capillary b) sample cone c) extraction cone d) 1st transfer hexapole and top-hat lens e) quadrupole mass analyser and transfer optics (two hexapoles) f) collision cell (filled with ~10^{-4} torr argon) g) transfer optics (hexapole) h) ToF entrance lenses i) ToF pusher assembly j) ToF drift region k) reflectron l) MCP detector. Arrows indicate the presence of turbomolecular pumps, which maintain the pressures close to the values shown. The drawing is not to scale.

2.3.3.1 MoQToF Topology

The addition of the drift tube was the major rearrangement imparted on the QToF-1 instrument. A schematic is shown in Figure 2.7. The drift tube was situated immediately after the first top hat lens (TH1) and is succeeded by a new hexapole (H2). Thus the resulting sections of the MoQToF are:

1. Ion source
2. Ion storage and transfer
3. Low-field drift tube
4. 1^{st} mass analyser (quadrupole)
5. Collision cell
6. 2^{nd} mass analyser (ToF)
7. Detector (MCPs)
8. Computerised data accumulation

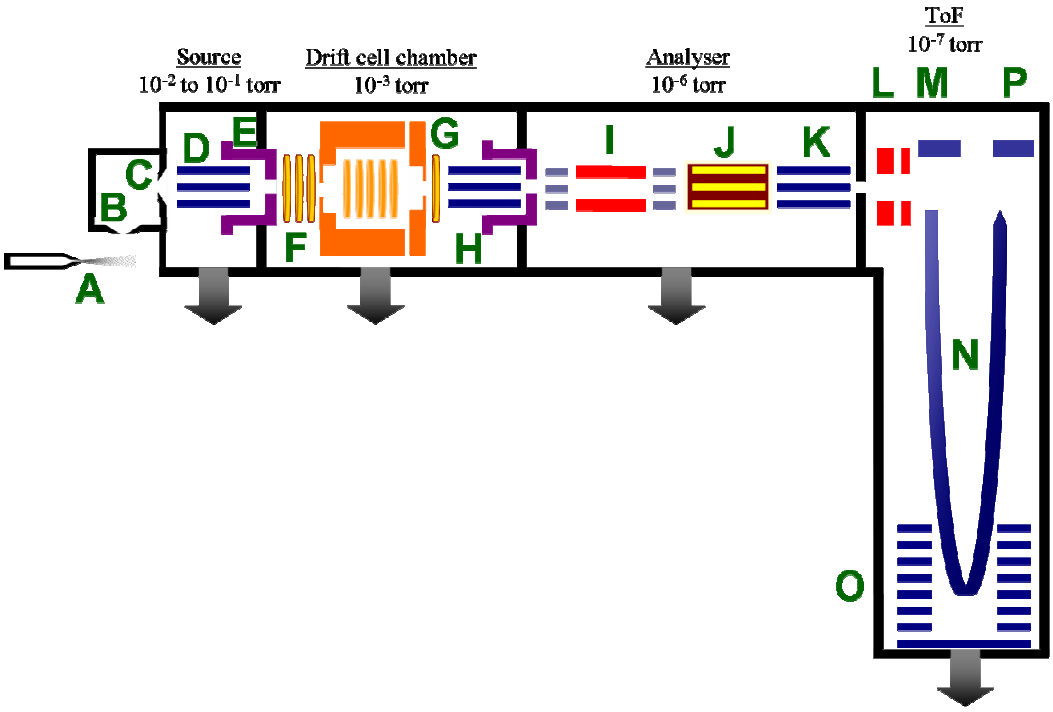


Figure 2.7 Schematic of a micromass Q-ToF. a) nESI capillary b) sample cone c) extraction cone d) 1st transfer hexapole [H1] where ion storage occurs e) 1st top hat lens [TH1] which aids trapping and pulsing of ions into the drift region f) Einzel lens [L1, L2, L3] g) drift tube [C1, C2] (filled with He at a pressure of 2.9-3.7 torr) and exit lens [L4] h) 2nd hexapole [H2] and 2nd top hat lens [TH2] transmitting ions into i) quadrupole mass analyser and transfer optics (two hexapoles) j) collision cell (filled with ~10^{-4} torr argon) k) transfer optics (hexapole) l) ToF entrance lenses m) ToF pusher assembly n) ToF drift region o) reflectron p) MCP detector. Arrows indicate the presence of turbomolecular pumps, which maintain the pressures close to the values shown. The schematic is not to scale.

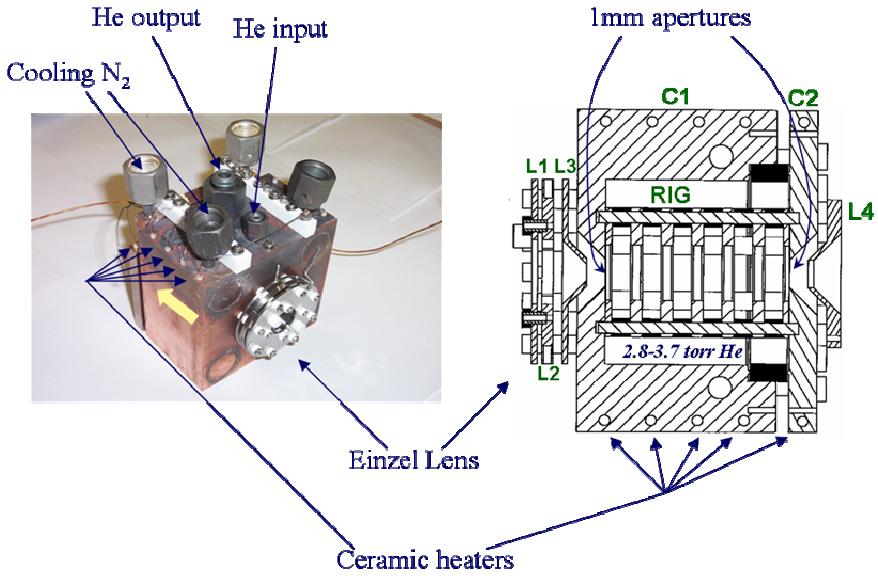


Figure 2.8 Drift cell photograph and alterior cross-sectional schematic. L1, L2, L3: Einzel lens. C1: cell body. C2: end cap. RIG: ring ion guide stack. Gas inlets and outlets are also shown.

2.3.3.2 Drift Tube Architecture

The drift cell assembly containing the region where ions drift is shown in Figure 2.8. As ions leave TH1, they encounter an Einzel lens which focuses and steers the beam into the entrance aperture. The Einzel lens itself consists of three cylindrical lenses (L1, L2, L3); L2 is split into four elements at the *xy* plane (x±, y±), a variable resistor interceding the power supply and the elements of each axis. Such a design is advantageous in that is allows the beam to be deflected accordingly and "correct" for a slightly off-axis positioning of the drift cell.

The drift tube itself is suspended from a supporting metal flange (where all appropriate electrical connections and drift gas lines are situated) by four metal rods. The drift tube consists of a cuboid copper block [cell body, C1] (x = y = 8.89cm, z=5.44cm) enclosing a cylindrical cavity (radius = 2.54cm (2.985cm where the ceramic seal is positioned), height = 4.674cm) in which ion drift takes place. The assembly is completed by and end cap (C2) which is connected to C1 by a ceramic seal. Six ceramic rods provide the base for mounding six equidistant copper drift rings (OD=3.556cm, ID=1.524cm, l=0.325cm). C1, each of the drift rings and C2 are each electrically connected to the next

by a 1MΩ resistor, thus distributing evenly the potential difference between C1 and C2, producing the constantly decreasing electric field required for IMS. The drift cell has removable apertures; all the experiments described here were performed with 1mm front and back apertures. The total drift length (distance between entrance and exit apertures) is 5.1cm. Exiting ions are focused by the tapering L4 element. Subsequently they are transferred to the quadrupole chamber by H2 and TH2.

2.3.3.3 MoQToF Operation

In IMMS mode, ions are produced by nESI as usual, and reach H1. A repulsive potential is applied onto TH1 allowing ions to be trapped in the RF region, since they are confined by the RF and cannot exit by either the extraction cone or TH1 since they are both kept at a higher voltage. At regular intervals (triggered by a Stanford DG535 pulse generator) the potential on TH1 is dropped to a value that allows transmission; thus, ideally, a square ion pulse is produced with width in the time domain equal to that of the pulse on TH1. The input pulse was kept between 20 and 45μs (this variance does not impede inter-experiment comparisons since measured ATDs are normally more than 150μs wide and space-charge effects are of minor significance at such ion densities) and was kept constant during every single mobility measurement. Upon the pulse onset the experimental time starts to be measured, and continues for a time equal to $200 \cdot t_{pusher}$; thus "time" in the current configuration is measured in pusher pulse periods, t_{pusher}.

Pulsed ions enter the cell where they drift and diffuse. It follows that ions are stored in H1 for up to this time period; yet there is no reason to assume that the dependence between the pulsed ion population and the original "gas-phase lifetime" is linear.

The drift region is filled with helium, at pressures between 2.9 and 3.7 torr. The pressure is measured by a capacitance manometer (MKS).

The injection voltage (potential difference between hexapole 1 and the cell body) was kept between 24 and 36V for all experiments described herein.

2.3.3.4 Temperature Regulation

The temperature of the drift cell is measured by three thermocouples; one is situated on C1 and two on C2 (top and bottom). The buffer gas (helium) will be heated by convection; thus the ion temperature is known since at low drift fields $T_{He} \approx T_{ion}$.

The drift cell is studded with resistive heaters (eight for the cell body and two for the end cap) which are assembled by threading 9cm-long alumina rods (each with 4 bores) with tantalum (with a resistivity of 135nΩm at 300K and a temperature coefficient of $0.0038K^{-1}$) wire (0.25mm in diameter, Goodfellow). The heaters are connected in series. A current is passed between the chain of heaters which, due to the resistance and current passed is converted to heat, as described by Joule's first law (*Q* here being heat, *I* current, *t* the heating time interval and *V* the voltage across the heaters):

$$Q = I^2 Rt = \frac{V^2}{R} t \qquad \{2.3\}$$

In turn, the ceramic elements, the copper cell housing the drift region and the drift gas are thus heated. Three remarks of practical significance can be made here. First, due to heat dissipation (caused among other reasons by the flow of gas through the cell as well as the existence of heat-conducting elements linking the cell to the instrument) the target temperature and the current supplied are proportional (in order to achieve $\dot{Q}_{dissipated} = \dot{Q}_{supplied}$. Secondly, heating ensued even after switching off the power supply, indicating that the heating rate was higher than the temperature equilibration rate. Thirdly, it quickly became evident that in order to retain signal, the pressure in the cell must be reduced; this observation can be explained due to increased diffusion, yielding broader peaks with a lower signal-to-noise ratio. Nevertheless, even at elevated temperatures, the field strength never exceeded that of 16Td, normally fluctuating between 5 and 12Td.

2.3.3.5 Estimation of Ionic Mobilities from IMMS Data

Momentum transfer theory and kinetic theory agree on the general form for the ionic mobility at low fields (see also the discussion in Appendix B). The more accurate result from kinetic theory is given in {1.23}.

We shall now proceed with a practical example for the analysis of an IMMS experiment of Bradykinin.

2.3.3.6 Analysis of Experimental Data

When the MoQToF is used in ion mobility mode, the data we obtain for a species of a particular m/z at a specific V_D looks like the chromatogram displayed in Figure 2.9.

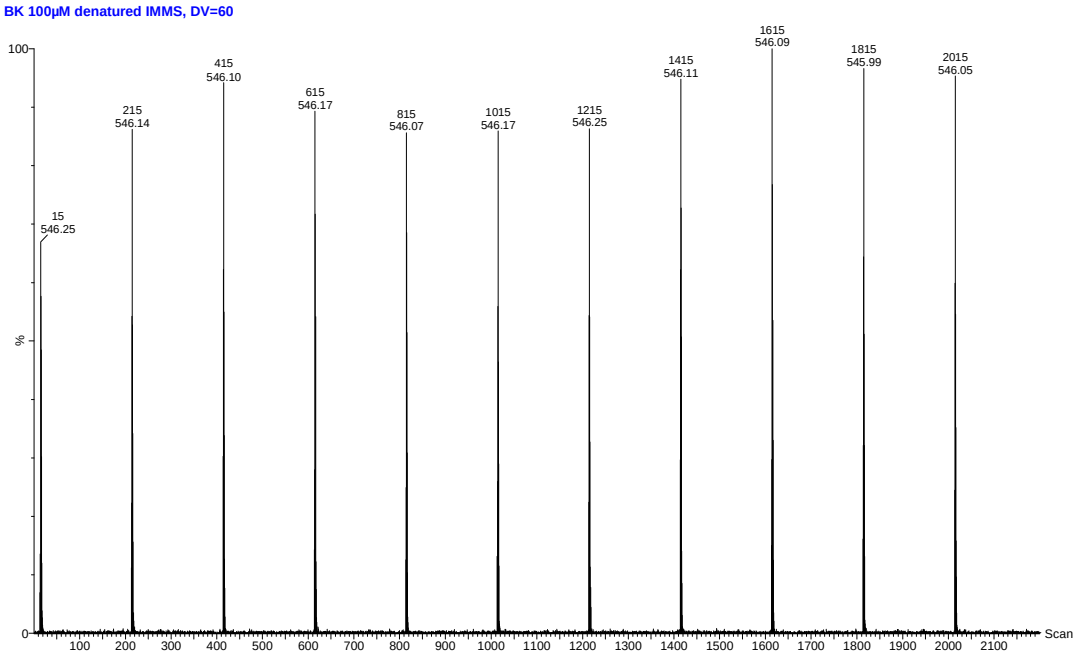


Figure 2.9 *masslynx* data for the 2+ charge state of bradykinin at a drift voltage of 50V

Essentially, this represents several repetitions of an ion mobility measurement, each having a duration of *ca*. 25 seconds. The maximum drift time recorded is 200 scans, each scan having the duration of one ToF pusher period (which is of the order of 100µs). Some of the inherent benefits of this method are that the ion intensity can be monitored during the experiment and occasionally the convergence of the experimental arrival time

distribution (ATD) can be monitored. Datasets like that in Figure 2.9 must be cropped and summed to obtain an average ATD like the one in Figure 2.10.

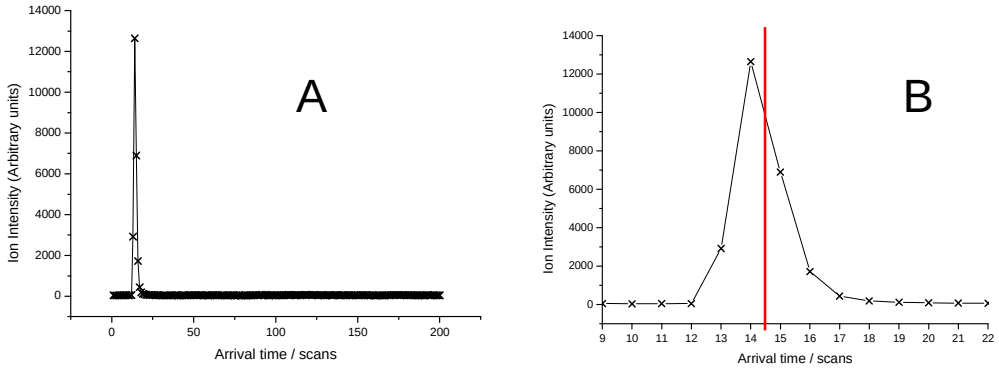


Figure 2.10 Average ATD obtained for BK^{2+} at 60V. B; zoomed in view of A – the points have been connected by lines in order to guide the eye. The red line shows the average scan obtained, 14.437 scans = 1010.59μs. This ATD was summed from 10 identical experiments each lasting 25 seconds with a pusher period of 70μs; i.e. it corresponds to $10\lfloor(25/7\cdot10^{-5})-1\rfloor$ = 3,571,410 ion pusles.

Assuming that the most significant features of the ATD exist between the detector's range, the ion intensity can be easily conceived as being directly proportional to the population of ion species that exit the drift tube and are transferred successfully to the detector. For our analysis we seek an estimate of the average time. Such a quantity (as well as any other time-dependent function) can be calculated directly from the experimental ATD:

$$\bar{t}_{a,V_D} = \frac{\sum tI(t)}{\sum I(t)} \qquad \{2.4\}$$

This method will yield the *average* arrival time of all species included in a particular peak of the mass spectrum; if several partially resolved conformers and/or multimeric states of the form $[nM+nzH]^{nz+}$ are present, fitting the ATD to function {1.30} is necessary in order to obtain more meaningful data. For practical reasons it was assumed that longitudal and transverse diffusion coefficients are equal; furthermore, as experimental ATDs were found to be broader than expected from the Nernst-Einstein-

Townsend relation (equation {1.29}), diffusion coefficients and ion mobilities were fitted as independent parameters. This is equivalent to assuming that an additional gaussian spread occurs as ions diffuse in instrument regions other than the drift tube – or that the presence of multiple unresolved conformations also takes this form. In the absence of higher-resolution data and the overall simplicity of the model, this approximation is justified. If the background noise levels are appreciable, a background subtraction of the average ATD before the averaging / fitting procedure may be useful. ToF MS scans can be easily converted to time by multiplying the scan number by the ToF pusher time. Furthermore, scans where the ion signal intensity is similar to the noise levels may be avoided, in order to prevent spurious signal from influencing our estimate of the arrival time.

The above process may be repeated for all voltages measured, thus yielding a list of average arrival times vs. *P/V* (Table 2.5; Figure 2.11). As expected, the arrival time does increase with increasing *P/V* and the relationship is linear as far as we can see, indicating that we are still within the low-field regime. The intercept at the time axis (≈681μs) corresponds to the 'dead' time, i.e. the time the ions spend in the instrument if no drift had taken place, and the slope (6603.3torrV^{-1}μs^{-1}) can be used to calculate the mobility and cross section from {1.23} and {1.26}. These were found to be 4.691±0.040 cm^2/Vs and 227.4±2.0 Å^2 respectively.

Table 2.5
Average arrival times obtained for a single run of bradykinin $[M+2H]^{2+}$

P/torr	Drift Voltage	P/V	Arrival Time (μs)
2.994	50.1	0.05976	1077.48
3.006	40	0.07515	1180.73
3.015	35	0.08614	1243.61
3.014	30	0.10047	1341.62
3.015	25	0.1206	1479.98
3.001	20	0.15005	1672.30

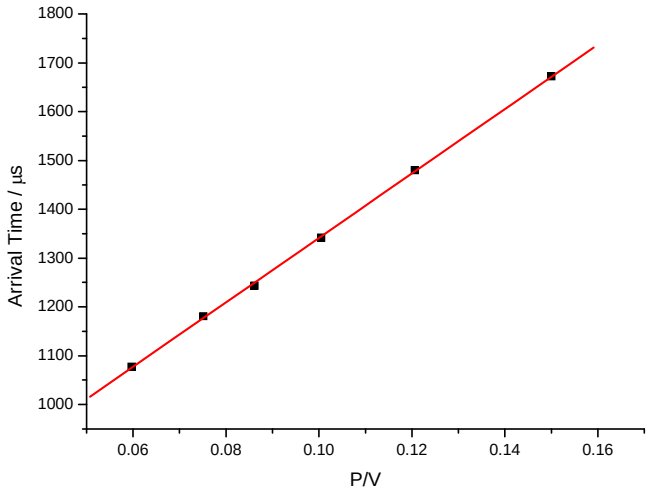


Figure 2.11. Plot of arrival time vs. P/V for the $[M+2H]^{2+}$ BK ion

2.3.4.1 Quadrupole Ion Trap

The other instrument that was used for mass spectrometry, CID and gas-phase HDX was a thermo-finnigan LCQ, an altogether different type of instrument compared to the ones with a Q-ToF architecture described above. A simple schematic of the instrument is shown in Figure 2.12. Like the other instruments, the LCQ was fitted with a nESI source which, in contrast to the Q-ToF instruments, is positioned along (or close to) the axis of a heated metallic capillary which transfers ions into the vacuum region. Subsequently two lenses, a skimmer preceeded by the tube lens, transfer the ions further into the instrument where they encounter two RF octapoles, separated by an interoctapole lens, which focus and inject the ion beam into the trap. Ions are thereby ejected from the trap to an electron multiplier where they are detected.

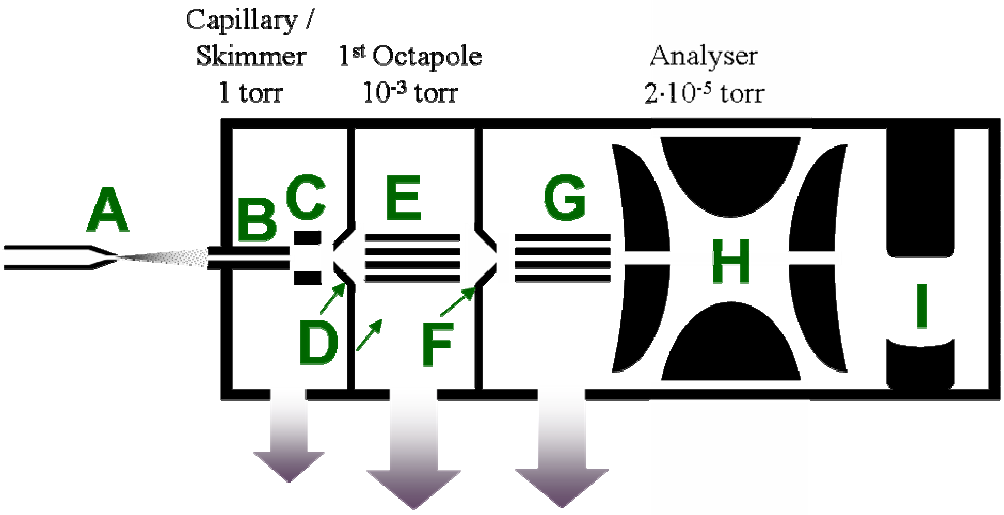


Figure 2.12 Schematic of the Finnigan LCQ. Three-stage differential pumping by a split-flow turbomolecular pump and an Edwards EM28 rotary pump maintain the pressures at the different regions close to the values shown. A pressure gauge is located in the analyser region. The diagram is not to scale. a) nESI tip b) heated capillary c) tube lens d) skimmer e) 1st octapole f) interoctapole lens g) 2nd octapole h) quadrupole ion trap (cavity filled with ~10^{-3}torr He) i) electron multiplier.

2.3.4.2 CID in a quadrupole ion trap

To perform CID with a Finnegan LCQ, ions of interest were mass-isolated (using an isolation window of 5-20amu) and subsequently activated by gradually ramping the RF potential applied on the end-cap electrode. Trajectories of ions in the QIT become marginally stable, resulting in longer trajectories and more energetic collisions. Eventually, collisions become sufficiently energetic to cause fragmentation of the peptide ion. The principles governing this process are summarised in section 1.5.8. Spectra were collected at a number of different normalised collision energies in order to generate the parent ion depletion curves presented in Chapter 3. The normalised collision energy is related to the applied RF potential by the following expression:

$$\text{Normalised collision energy} = a\frac{m_i \cdot RF_{slope} + RF_{intercept}}{RF} \qquad \{2.5\}$$

RF represents the amplitude of the applied alternating current, m_i the mass of the ion, RF_{slope} and $RF_{intercept}$ are instrument-specific variables derived from calibration processes

and the correction factor $a = 0.3$ is an approximation of the m/z of the 'parent' ion and the lowest m/z that can be maintained in the QIT under these operation conditions. A much more comprehensive description of the methodology used can be found elsewhere[22].

2.3.4.3 Gas-phase HDX

As ions can be stored in a quadrupole ion trap for specified periods of time, such a platform presents a setup favourable for implementing gas-phase HDX. HDX experiments in quadrupole ion trap have been previously reported: this setup yields data in reasonable agreement with HDX in an ICR cell[23]. The ion trap is filled with helium (to pressures of ~10^{-4} to 10^{-3} torr) which acts as a buffer gas to collisionally cool the ions and also to vibrationally activate them during CID. To perform gas-phase HDX, a reservoir containing deuterated solvent (typically d_4-methanol, MeOD) was interpolated between the helium inlet and the instrument; the setup is outlined in Figure 2.13. The trap can now contain MeOD in addition to He; collisions of analyte ions with MeOD can result in gas-phase HDX occurring. The MeOD reservoir can be by-passed in order to ensure that the flow of He/MeOD into the trap is similar to that of He alone. Given that subsequent experiments are performed at similar pressures of He and similar ambient conditions, the amount of observed exchange events as well as HDX kinetics can provide information about the relative conformational stability of biomolecular ions[23, 24].

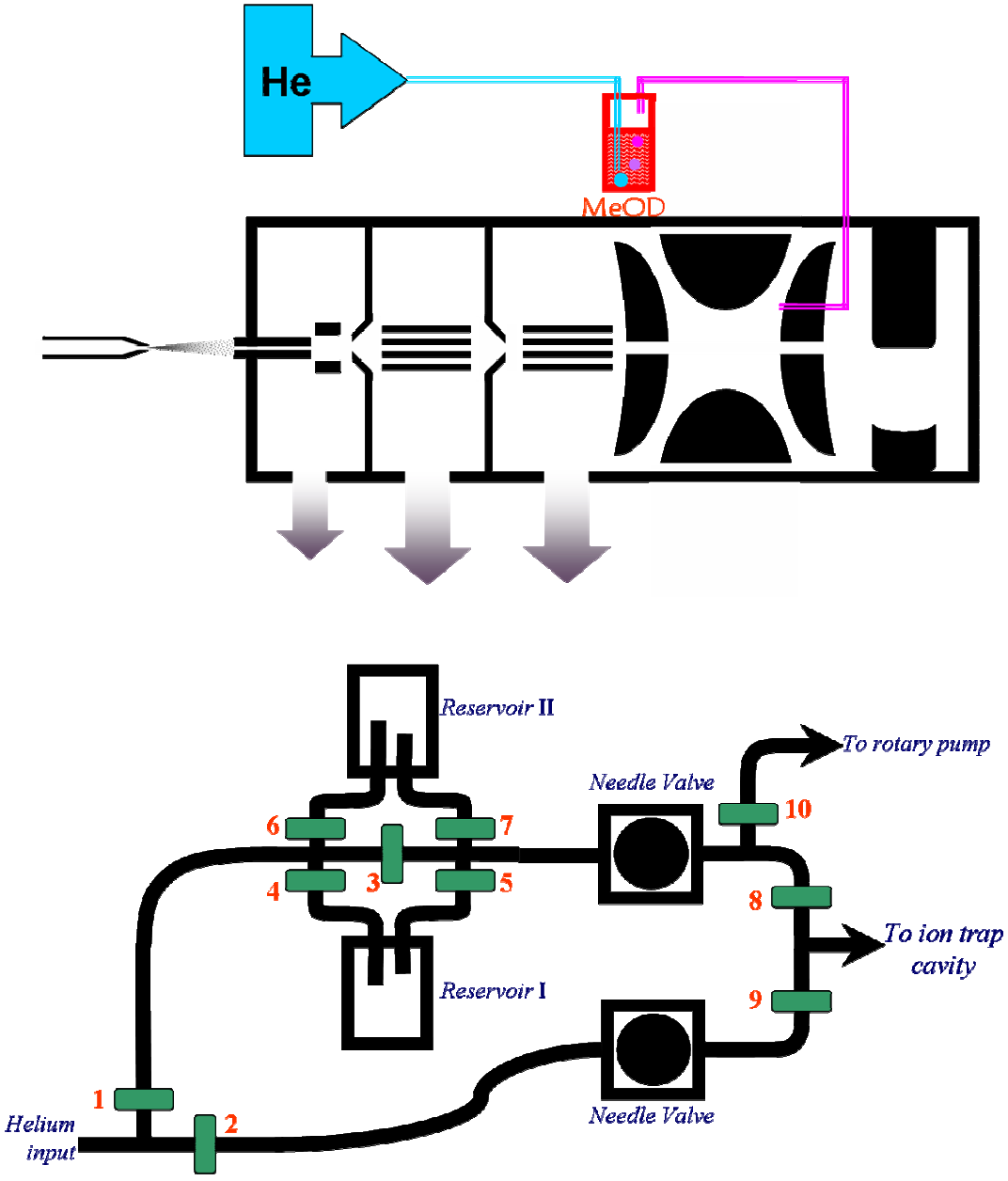


Figure 2.13 a) LCQ instrument schematic with He/MeOD inlet. Helium (20-60psi) is bubbled through a reservoir containing MeOD, thus carrying MeOD into the trap cavity. b) Outline of the He and He/MeOD inlet; valves shown in green. In order to adjust the He/MeOD needle valve the pressure in the analyser region is recorded with valves 2 and 9 open, all others being closed. Subsequently, valves 2 and 9 are closed and 1, 3 and 8 are opened. The needle valve is adjusted until the ion gauge reading is identical to the one recorded. He buffer gas can be filtered through the constituents of reservoir I or II by closing valve 3 and opening 4 and 5 or 6 and 7, respectively. The lines can be evacuated by opening valve 10 (4-7 must be closed)

2.4 Circular Dichroism

Circular dichroism (CD) spectra display the differential absorption of right-handed and left-handed circularly polarised light as a function of wavelength of the incident light beam. 2° and higher-order structural elements of peptides and proteins are chiral and their resulting CD spectra contain a wealth of information. Far-UV CD spectra (180-260nm, the region where the peptide bond absorbs) are used routinely to measure the content of secondary structure of peptides, whereas near-UV spectra (260-350nm) can reveal conformational strain of aromatics (Trp, Tyr, Phe) and even the presence of disulfide bidges[25]. CD spectra were collected on a Jasco J810 spectropolarimeter (courtesy of Prof. Lindsay Sawyer). Far-UV spectra were recorded at a 190-260nm range, for which a quartz cuvette (Hellma) with a path length of 1mm was used; near-UV spectra spanned 250 to 340nm, using a cuvette with 10mm path length. Typically each far-UV spectrum consists of an average of three consecutive measurements; an average of five measurements was employed for near-UV spectra. The temperature of the sample was regulated by means of a peltier temperature regulator. Typical values of the scan settings are displayed in Table 2.6.

Table 2.6
Typical ranges for CD scan settings

Parameter	Typical Value (far UV CD)	Typical Value (near UV CD)
Sensitivity	Standard	Standard
Wavelength range	190-260 nm	250-340 nm
Data pitch	0.2 or 0.5 nm	0.2 or 0.5 nm
Scanning mode	Continuous	Continuous
Scanning Speed	10 or 20 nm/min	10 or 20 nm/min
Response	2 s	2 s
Bandwidth	1 nm	1 nm
Accumulation	3	5

Fixed-wavelength temperature scans discussed in Chapter 4 were analysed using a simple two-state thermodynamic model. The entire peptide population is thought as undergoing two interconverting states, say A and B (analogous to ‘folded’ and ‘unfolded’ states):

$$\text{A} \quad \Leftrightarrow \quad \text{B}$$

At every temperature the total number of peptide population is sequestered between the two states, with proper normalisation (*P* being probability)

$$P_A + P_B = 1 \qquad \{2.6\}$$

Defining $\Delta G = G_B - G_A$, P_A and P_B can be expressed as

$$P_A = \frac{1}{1+e^{-\frac{\Delta G}{RT}}} \qquad \{2.7\}$$

$$P_B = \frac{e^{-\frac{\Delta G}{RT}}}{1+e^{-\frac{\Delta G}{RT}}}$$

As the spectroscopic properties of Trp cage at low concentrations is supposed to be additive, varying linearly with concentration, the detected ellipticity ϑ_{tot} is related to the population of the two species:

$$\frac{1}{c_0 \ell}\vartheta_{tot} = \vartheta_A P_A + \vartheta_B P_B = \frac{\vartheta_A + \vartheta_B e^{-\frac{\Delta G}{RT}}}{1+e^{-\frac{\Delta G}{RT}}} \qquad \{2.8\}$$

$c_0, \ell, \vartheta_A, \vartheta_B$ being the total peptide concentration, the cuvette path length and the molar ellipticities ("extinction coefficients") of species A and B, respectively. The sigmoidal unfolding curves can be fitted to expression {2.8} to obtain values for the enthalpy and entropy differences between the two states. To demonstrate graphically that the system behaves consistently with this simple model at a macroscopic level, a van't Hoff plot will be constructed:

$$\Delta G = \Delta H - T\Delta S = -RT\ln K \Rightarrow -R\ln K = \Delta H\frac{1}{T} - \Delta S \qquad \{2.9\}$$

Now, the equilibrium constant *K* can be expressed in terms of only $\vartheta_{tot}, \vartheta_A$ and ϑ_B:

$$K = \frac{P_B}{P_A}\frac{(\vartheta_B - \vartheta_A)}{(\vartheta_B - \vartheta_A)} = \frac{\vartheta_A(P_A - 1) + \vartheta_B P_B}{\vartheta_B(1 - P_B) - \vartheta_A P_A} = \frac{\vartheta_A P_A + \vartheta_B P_B - \vartheta_A}{\vartheta_B - \vartheta_A P_A - \vartheta_B P_B} = \frac{\vartheta_{tot} - c_0 \ell \vartheta_A}{c_0 \ell \vartheta_B - \vartheta_{tot}} \qquad \{2.10\}$$

To summarise, an initial fit of experimental data achieves estimation of ϑ_A and ϑ_B as well as ΔG; a subsequent plot of $-R\ln K$ against the inverse temperature (*cf* equation

{2.9}) indicates consistency with the model as well as confirming the values obtained for ΔG and ΔS.

2.5 Peptides and Proteins

Trp cage constructs, L_4PL_4K, TTR and defensins were synthesised using Fmoc solid phase peptide sythesis methods[26]. I am very grateful to Dr. Derek MacMillan (UCL) for sythesising the Trp cage, defensins and L_4PL_4K. TTR was kindly offered by Dr. Cait McPhee. Other peptides and proteins (bradykinin, melittin, egg white lysozyme) were purchased from Sigma-Aldrich.

2.6 Final Notes on Methodology

So far a strategy has been described to characterise the conformations of biomolecules in a gas-phase environment. IMMS, along with other MS-based techniques has been applied to probe the structure of peptide ions *in vacuo*; in some cases the structures of studied peptides in solution were monitored spectroscopically by CD; experimental data were recalled to validate structures generated by molecular mechanics based methods. The following chapters will illustrate the progress made using this strategy. It is now appropriate to mention the areas left obscure by this method and point some improvements which although unfortunately not pursued here in full, will certainly prove most beneficial upon their fulfilment. Dimensionality, as it has been mentioned already, is of crucial importance when considering peptide configurations. Representing a peptide by the estimated He collision cross-sections of the resolved conformers is equivalent to projecting the entire PES onto one (or, at best, a small set of) zero-dimensional points. However, more information can be extracted from collected datasets, which themselves occupy four dimensions:

- The m/z axis
- Drift time
- Drift voltage
- Intensity

It is clear that experimental arrival time distributions of proteins at room temperature are broader than those expected for single species[27]. Albeit the effect ambient conditions and instrumental parameters may have on the shape of the ATD, recalling the continuity equation:

$$\frac{\partial J_i}{\partial t} = \nabla D_i \bullet \nabla n_i - v_D \bullet \nabla n_i + \sum_j \left(\alpha_{j \to i} n_j - \alpha_{i \to j} n_i \right) \quad \{2.11\}$$

shows that the appearance of the arrival time spectrum will depend on the conformational stability and dynamics of the analyte. The last part of the above expression is in essence a rendering of the master equation. The difficulty here lies in knowing the transition rate constants $\alpha_{j \to i}$ and $\alpha_{i \to j}$, save a relevant initial distribution over all states. Such problems are brought about by the sheer number of states that a peptide system can possible occupy; involved solutions suggested for example by Gatland[28] may fail to capture the entire scope of possible transitions. Nevertheless, Becker and Karplus[29] have demonstrated that master equation dynamics analysis can be applied to peptides after a 'coarse-graining' of the energy dimension, thus revealing the broader features of the PES, without focusing on minor transitions. Furthermore, a study by Komatsuzaki *et al.*[30] has shown that the general topological features of energy landscapes can be adequately represented by models with some degree of coarse-graining. With such methods at hand, even the demanding treatment required for the solution of eq. {2.11} may become a viable approach for understanding the structures of large conformationally variant systems from ion mobility data. High-resolution IMMS instruments[31] as well as tandem-IMS devices[32] will provide less ambiguous data, enabling even more solid correspondences between theory and experiment to be made. An intriguing field for testing and finally employing such methods would be molecular conformational switches, which undergo significant conformational changes upon a known perturbation. Numerous systems displaying such properties have been designed in recent years and many of these are novel biomimetic compounds[33], or biomolecules with desirable properties[34]. It is thus very likely that the ongoing swelling of computational and experimental technologies will in the future promote the growth of numerous fields of research.

2.7 References

1. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *Journal of Computational Chemistry* **2005,** *26* (16), 1668-1688; Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P., Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Computer Physics Communications* **1995,** *91* (1-3), 1-41.
2. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *Journal of the American Chemical Society* **1995,** *117* (19), 5179-5197.
3. Wang, J. M.; Cieplak, P.; Kollman, P. A., How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry* **2000,** *21* (12), 1049-1074.
4. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A., A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *Journal of Physical Chemistry* **1993,** *97* (40), 10269-10280; Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A., Application of Resp Charges to Calculate Conformational Energies, Hydrogen-Bond Energies, and Free-Energies of Solvation. *Journal of the American Chemical Society* **1993,** *115* (21), 9620-9631.
5. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R., Molecular-Dynamics with Coupling to an External Bath. *Journal of Chemical Physics* **1984,** *81* (8), 3684-3690.
6. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., NUMERICAL-INTEGRATION OF CARTESIAN EQUATIONS OF MOTION OF A SYSTEM WITH CONSTRAINTS - MOLECULAR-DYNAMICS OF N-ALKANES. *Journal of Computational Physics* **1977,** *23* (3), 327-341.
7. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G., Pairwise Solute Descreening of Solute Charges from a Dielectric Medium. *Chemical Physics Letters* **1995,** *246* (1-2), 122-129; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G., Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *Journal of Physical Chemistry* **1996,** *100* (51), 19824-19839.
8. Tsui, V.; Case, D. A., Theory and applications of the generalized Born solvation model in macromolecular Simulations. *Biopolymers* **2000,** *56* (4), 275-291.
9. Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P., OPTIMIZATION BY SIMULATED ANNEALING. *Science* **1983,** *220* (4598), 671-680.
10. Randelman, R. E.; Grest, G. S., N-City Traveling Salesman Problem - Optimization by Simulated Annealings. *Journal of Statistical Physics* **1986,** *45* (5-6), 885-890.
11. Wales, D. J., *Energy landscapes*. Cambridge University Press: Cambridge, UK ; New York, 2003; p ix, 681 p.
12. Grest, G. S.; Soukoulis, C. M.; Levin, K., Cooling-Rate Dependence for the Spin-Glass Ground-State Energy - Implications for Optimization by Simulated Annealing. *Physical Review Letters* **1986,** *56* (11), 1148-1151.
13. Huse, D. A.; Fisher, D. S., Residual Energies after Slow Cooling of Disordered-Systems. *Physical Review Letters* **1986,** *57* (17), 2203-2206.
14. Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F., Structural information from ion mobility measurements: Effects of the long-range potential. *Journal of Physical Chemistry* **1996,** *100* (40), 16082-16086; Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F., Structural information from ion mobility measurements: Effects of the long-range potential (vol 100, pg 16082, 1996). *Journal of Physical Chemistry A* **1997,** *101* (5), 968-968.
15. Shvartsburg, A. A.; Jarrold, M. F., An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chemical Physics Letters* **1996,** *261* (1-2), 86-91.
16. Mack, E., Average cross-sectional areas of molecules by gaseous diffusion methods. *Journal of the American Chemical Society* **1925,** *47*, 2468-2482.

17. Vonhelden, G.; Hsu, M. T.; Gotts, N.; Bowers, M. T., Carbon Cluster Cations with up to 84 Atoms - Structures, Formation Mechanism, and Reactivity. *Journal of Physical Chemistry* **1993,** *97* (31), 8182-8192; Wyttenbach, T.; vonHelden, G.; Batka, J. J.; Carlat, D.; Bowers, M. T., Effect of the long-range potential on ion mobility measurements. *Journal of the American Society for Mass Spectrometry* **1997,** *8* (3), 275-282.
18. McCullough, B. J.; Kalapothakis, J.; Eastwood, H.; Kemper, P.; MacMillan, D.; Taylor, K.; Dorin, J.; Barran, P. E., Development of an ion mobility quadrupole time of flight mass spectrometer. *Analytical Chemistry* **2008,** *80* (16), 6336-6344; McCullough, B. J. Development of an ion mobility mass spectrometer to study gas phase conformations of biomolecules. Thesis (Ph D), University of Edinburgh, 2007.
19. MacPhee, C., 2008.
20. Yamashita, M.; Fenn, J. B., ELECTROSPRAY ION-SOURCE - ANOTHER VARIATION ON THE FREE-JET THEME. *Journal of Physical Chemistry* **1984,** *88* (20), 4451-4459.
21. Hoffmann, E. d.; Stroobant, V., *Mass spectrometry : principles and applications*. 3rd ed.; J. Wiley: Chichester, West Sussex, England ; Hoboken, NJ, 2007.
22. Stopford, A. P. Structural and thermodynamic analysis of peptide and protein ions in solution and in the gas phase. Thesis (Ph D), University of Edinburgh, 2009.
23. Evans, S. E.; Lueck, N.; Marzluff, E. M., Gas phase hydrogen/deuterium exchange of proteins in an ion trap mass spectrometer. *International Journal of Mass Spectrometry* **2003,** *222* (1-3), 175-187.
24. Wagner, D. S.; Anderegg, R. J., CONFORMATION OR CYTOCHROME-C STUDIED BY DEUTERIUM-EXCHANGE ELECTROSPRAY-IONIZATION MASS-SPECTROMETRY. *Analytical Chemistry* **1994,** *66* (5), 706-711; Campbell, S.; Rodgers, M. T.; Marzluff, E. M.; Beauchamp, J. L., Deuterium exchange reactions as a probe of biomolecule structure. Fundamental studies of cas phase H/D exchange reactions of protonated glycine oligomers with D2O, CD3OD, CD3CO2D, and ND3. *Journal of the American Chemical Society* **1995,** *117* (51), 12840-12854; Purves, R. W.; Barnett, D. A.; Ells, B.; Guevremont, R., Gas-phase conformers of the [M+2H](2+) ion of bradykinin investigated by combining high-field asymmetric waveform ion mobility spectrometry, hydrogen/deuterium exchange, and energy-loss measurements. *Rapid Communications in Mass Spectrometry* **2001,** *15* (16), 1453-1456; Valentine, S. J.; Clemmer, D. E., H/D exchange levels of shape-resolved cytochrome c conformers in the gas phase. *Journal of the American Chemical Society* **1997,** *119* (15), 3558-3566; Valentine, S. J.; Clemmer, D. E., Temperature-dependent H/D exchange of compact and elongated cytochrome c ions in the gas phase. *Journal of the American Society for Mass Spectrometry* **2002,** *13* (5), 506-517.
25. Kelly, S. M.; Price, N. C., The application of circular dichroism to studies of protein folding and unfolding. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology* **1997,** *1338* (2), 161-185.
26. Chan, W. C.; White, P. D., *Fmoc solid phase peptide synthesis : a practical approach*. Oxford University Press: New York, 2000; p xxiv, 346 p.
27. Wyttenbach, T.; vonHelden, G.; Bowers, M. T., Gas-phase conformation of biological molecules: Bradykinin. *Journal of the American Chemical Society* **1996,** *118* (35), 8355-8364; Shelimov, K. B.; Jarrold, M. F., Conformations, unfolding, and refolding of apomyoglobin in vacuum: An activation barrier for gas-phase protein folding. *Journal of the American Chemical Society* **1997,** *119* (13), 2987-2994; Koeniger, S. L.; Merenbloom, S. I.; Clemmer, D. E., Evidence for many resolvable structures within conformation types of electrosprayed ubiquitin ions. *Journal of Physical Chemistry B* **2006,** *110* (13), 7017-7021; Koeniger, S. L.; Merenbloom, S. I.; Sevugarajan, S.; Clemmer, D. E., Transfer of structural elements from compact to extended states in unsolvated ubiquitin. *Journal of the American Chemical Society* **2006,** *128* (35), 11713-11719; Koeniger, S. L.; Clemmer, D. E., Resolution and structural transitions of elongated states of ubiquitin. *Journal of the American Society for Mass Spectrometry* **2007,** *18* (2), 322-331.
28. R., G. I., Analysis for Ion Drift Tube Experiments. *Case Studies in Atomic Physics* **1974,** *4*, 369-437.
29. Becker, O. M.; Karplus, M., The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *Journal of Chemical Physics* **1997,** *106* (4), 1495-1517.

30. Komatsuzaki, T.; Hoshino, K.; Matsunaga, Y.; Rylance, G. J.; Johnston, R. L.; Wales, D. J., How many dimensions are required to approximate the potential energy landscape of a model protein? *Journal of Chemical Physics* **2005,** *122* (8).
31. Dugourd, P.; Hudgins, R. R.; Clemmer, D. E.; Jarrold, M. F., High-resolution ion mobility measurements. *Review of Scientific Instruments* **1997,** *68* (2), 1122-1129; Tang, K.; Shvartsburg, A. A.; Lee, H. N.; Prior, D. C.; Buschbach, M. A.; Li, F. M.; Tolmachev, A. V.; Anderson, G. A.; Smith, R. D., High-sensitivity ion mobility spectrometry/mass spectrometry using electrodynamic ion funnel interfaces. *Analytical Chemistry* **2005,** *77* (10), 3330-3339; Kemper, P. R.; Dupuis, N. F.; Bowers, M. T., A new, higher resolution, ion mobility mass spectrometer. *International Journal of Mass Spectrometry* **2009**, Article in press.
32. Koeniger, S. L.; Merenbloom, S. I.; Valentine, S. J.; Jarrold, M. F.; Udseth, H. R.; Smith, R. D.; Clemmer, D. E., An IMS-IMS analogue of MS-MS. *Anal Chem* **2006,** *78* (12), 4161-74; Merenbloom, S. I.; Koeniger, S. L.; Bohrer, B. C.; Valentine, S. J.; Clemmer, D. E., Improving the efficiency of IMS-IMS by a combing technique. *Analytical Chemistry* **2008,** *80* (6), 1918-1927; Merenbloom, S. I.; Koeniger, S. L.; Valentine, S. J.; Plasencia, M. D.; Clemmer, D. E., IMS-IMS and IMS-IMS-IMS/MS for separating peptide and protein fragment ions. *Analytical Chemistry* **2006,** *78* (8), 2802-2809.
33. Prince, R. B.; Okada, T.; Moore, J. S., Controlling the secondary structure of nonbiological oligomers with solvophobic and coordination interactions. *Angewandte Chemie-International Edition* **1999,** *38* (1-2), 233-236.
34. Kinnear, B. S.; Hartings, M. R.; Jarrold, M. F., Helix unfolding in unsolvated peptides. *Journal of the American Chemical Society* **2001,** *123* (24), 5660-5667; Sudha, R.; Jarrold, M. F., Left-handed and ambidextrous helices in the gas phase. *Journal of Physical Chemistry B* **2005,** *109* (23), 11777-11780; Cerasoli, E.; Sharpe, B. K.; Woolfson, D. N., ZiCo: A peptide designed to switch folded state upon binding zinc. *Journal of the American Chemical Society* **2005,** *127* (43), 15008-15009; Channon, K.; MacPhee, C. E., Possibilities for 'smart' materials exploiting the self-assembly of polypeptides into fibrils. *Soft Matter* **2008,** *4* (4), 647-652; Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D., Design of a novel globular protein fold with atomic-level accuracy. *Science* **2003,** *302* (5649), 1364-1368.

3

Conformational traits of the Trp cage miniprotein in the gas phase at ambient temperature

3.1 Abstract

Having revealed the details of our research methodology in the previous chapter we now apply it to the Trp cage miniprotein. In agreement with solution studies, MS, IMMS and gas-phase HDX experiments point univocally to the compactness of the gas-phase structure of this synthetic peptide. Molecular mechanics simulations reveal the preservation of helical elements albeit with partial loss of the tertiary structure. Simulations also predict a major role for charge location on desolvated protein conformation; to further investigate the effect of charge location a number of variant sequences were synthesised (by replacing/swapping arginine and lysine residues or engineering protecting groups at the termini) and examined by the same procedures. A variant containing *d*-proline at position 12 was also investigated. All peptide ions are found to be compactly folded *in vacuo*, posing intriguing challenges for the characterisation of native protein structures by mass spectrometry around ambient temperature.

3.2 The Trp cage Miniprotein

3.2.1 The Origins of Trp Cage

Trp cage is the product of a protein design effort; its ancestry can be traced back to extendin-4 (a 39-amino acid membrane-interacting helical peptide originally isolated from the Gila monster (the lizard *Heloderma suspectum*) saliva) and the research of the Andersen group[2, 3].

Interest in extendin-4 was initially triggered by its potential as a novel therapeutic agent for diabetes; the aqueous and micelle-bound structures of the peptide revealed the formation of a (weakly preserved) globular fold at its C-terminus, held together by a hydrophobic core when solvated by water. Truncation and subsequent mutations of the peptide optimized the fold: an N-terminal asparagine was added to 'cap' the helix, A35 was mutated to an R and N28 to D in order to promote a K27-D28-R35 salt bridge; for this purpose E24 was changed to Q to minimize disruption of this feature. Finally, an F22Y mutation was introduced in order to reinforce π-π stacking with W25. Following such modifications, at least 95% of the peptide population was found to be folded in aqueous conditions (the peptide being even more stable in water containing small amounts of 2,2,2-trifluoroethanol – TFE)[2].

Table 3.1
Sequences of Trp cage and the parental peptide extendin-4

	1	**11**	**21**	**31**	**39**
Extendin-4	**HGEGTFTSDL**	**SKQMEEEAVR**	**LFIEWLKNGG**	**PSSGAPPPS**	
Trp cage			N **LY**I**QWLK**D**GG**	**PSSG**R**PPPS**	
			1	**11**	**20**

Figure 3.1
Heloderma suspectum, or Gila monster.

3.2.2 Structure of the Trp Cage

The Trp cage miniprotein encompasses two astonishing traits that has rendered it an important test system for both biomolecular simulations and new experimental techniques, namely

- The way it "encodes" a distinct tertiary structure in just 20 amino acids
- Its fast folding rate ($k_{fold} = 4.1\mu s$) and the apparent stability of its fold.

The aqueous fold of the Trp cage forms (starting from the N-terminus) an α-helix followed by a short loop, then one 3_{10} helical turn, another loop and a PPII ribbon (Figure 3.2). Despite its small size the peptide contains a hydrophobic core which includes Y3, W6, L7, P12, P17, P18 and P19. The key interactions that hold the fold together in an aqueous environment (and to which the peptide owes its name) are the CH – π contacts between the aliphatic proline side-chains and the aromatic indole ring in W6, causing the 3_{10} helical turn and the C-terminal PPII 'helix' to arrange in close proximity to the central tryptophan. As discussed by Neidigh *et al.*[2], the K8-D9-R16 salt bridge is significant in stabilizing the tertiary fold[4] and – as it will be further discussed

later – it is even more important for the gas-phase structure. Other traits, such as the helix-stabilising N1 and the Q5-D9 contact also contribute to the stability of the fold[2].

Not only is Trp cage the smallest stable miniprotein containing tertiary contacts, it is the fastest protein that folds cooperatively as well[5]. Fluorescence and CD measurements of thermal unfolding/refolding have elucidated a folding rate constant $k_f \approx 2.4 \cdot 10^5$ s^{-1} (folding time ≈4μs), approaching the limit posed by Brownian diffusion (~10^6 s^{-1})[5]. A possible reason for this might be the retention of residual "native-like" contacts even at high temperatures[4, 6, 7]. Neidigh *et al.*[2] suggest these may be a W6-P12 hydrophobic cluster, although theoretical data tend to present the D9-R16 ionic interaction as more prominent[4].

These features paved the way for the establishment of Trp cage as a widespread model for protein folding.

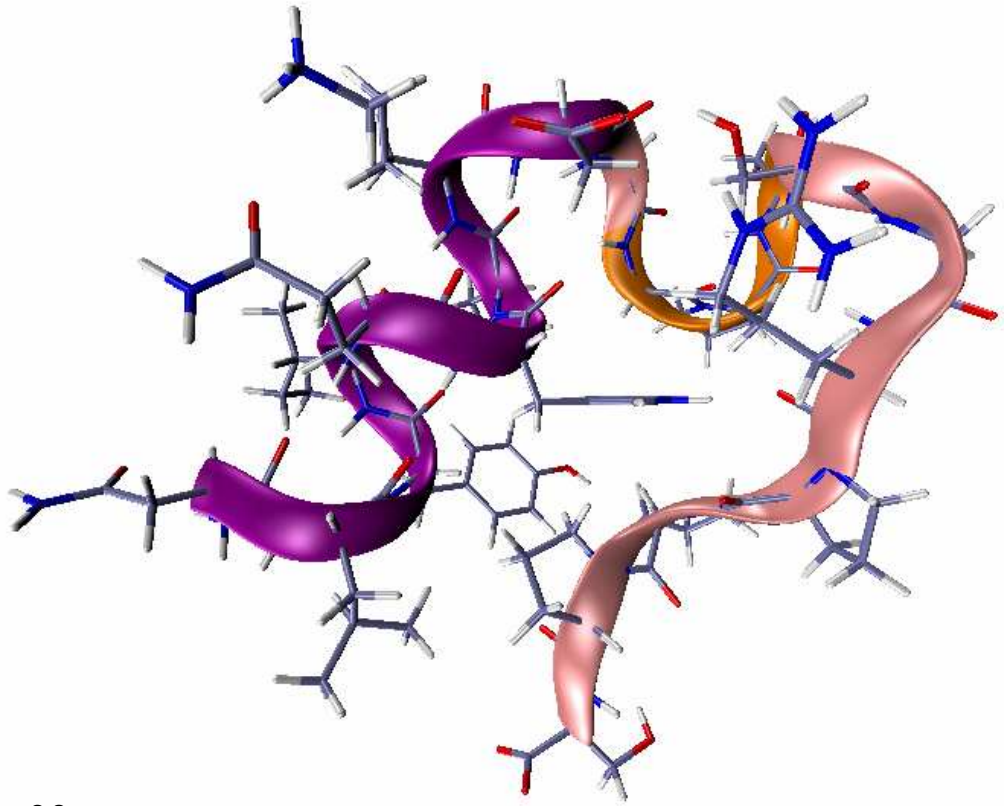


Figure 3.2
The Trp cage fold (pdb code 1L2Y) in an aqueous environment. The folded peptide contains an N-terminal α-helix (purple), a single 3_{10} helical turn (orange) – which can be alternatively described as three consecutive type-II turns – and, finally, a PPII helix (pink). It has a tightly packed hydrophobic core stabilised primarily by proline (aliphatic) to tryptophan (aromatic) interactions. Image produced with VMD[1].

3.2.3 The Trp Cage "Legacy"

The design of the Trp cage was followed by a flood of publications by both theorists and experimentalists, establishing this peptide as an important benchmark for both protein simulations and novel experimental methodologies[5, 8].

3.2.3.1 Computational Studies

The first study predicting the all-atom structure with high accuracy for the optimized peptide was published only a few months after its first appearance in the literature[9] and numerous other studies followed[4, 8, 10-13]. The ease with which its 'native' structure was predicted is another characteristic of the Trp cage that along with its small size, stability and unusually fast folding rate render it atypical of most proteins. The methods used to investigate its structure computationally range from MD simulations to much more advanced methods aimed to sample conformational space more comprehensively, such as replica exchange[4, 6, 10] and other stochastic methods[8, 10, 11]. The findings from these studies are intriguing. Molecular dynamics simulations of the Trp cage have provided detailed insights into its folding mechanism and even allowed some comparisons with experimental data to be made [4, 7, 13-15]. In the earliest of these studies, Simmerling *et al.*[9] predicted with great accuracy the structure of the peptide in a solvent continuum using the AMBER99 force field. Longer MD runs[12] using the "improved" Duan *et al.*[16] force field corroborated the earlier results. The study of the "unfolded" ensemble at a molecular level is an interesting result of the MD simulations; in agreement with the NMR results, unfolded conformations retaining the hydrophobic contact W6-P12 were found providing a possible explanation for the extremely fast folding time of the protein. Although a small number of MD simulations cannot provide accurate estimates of folding times, use of the folding@home distributed computing project (led by V. J. Pande) allowed several thousands of independent short MD simulations to be run in parallel, on the basis of which a reasonable estimate of the folding rate constant[13] could be determined. The estimate depended on the definition of a "folded structure", determined by an α-carbon root-mean-square deviation ($RMSD_{C\alpha}$) cut-off value. With

$RMSD_{C\alpha}$ = 3.0 Å, τ = 1.5 μs and when $RMSD_{C\alpha}$ = 2.5 Å, τ = 8.7 μs (experimental τ = 4.1 μs)[13].

The thermodynamic properties of the peptide have been investigated using replica-exchange methods[4,6]; these studies are reviewed in Chapter 4. Although the Trp cage is not as well studied in the gas phase as it is in solution, simulations of the peptide in a solvent-free environment have been carried out. A replica-exchange methodology has been adopted for the investigation of the gas-phase structure of Trp cage and the results have been used to interpret ECD fragmentation data, as it will be discussed later[15].

Once such strong links between theory and experiment were established, the Trp cage was accepted as an important benchmark for conformational searching algorithms[8, 10, 11, 13], but also for testing of promising experimental methods, which are undoubtedly rendered more powerful when used in conjunction with molecular modelling.

3.2.3.2 Experimental Studies

Although the Trp cage was designed in order to facilitate testing of simulation methods applied to protein-like systems, its role for testing experimental methods should not be overlooked. Soon after the first appearance of the Trp cage miniprotein, a study reporting on the kinetics of the folding process by CD and fluorescence spectroscopy was published[5]. The peptide contains both an α-helical segment, which renders it readily detectable by CD, and one structurally important Trp residue, allowing fluorescence studies to be performed. CD and fluorescence results were in excellent agreement and provided kinetic and thermodynamic parameters that could be compared to theoretically derived quantities, as discussed above, assuming a two-state folding mechanism. However, Ahmed *et al.*[7] have gone a step beyond simple two-state folding models even for this optimized sequence by attempting to capture the 'multi-basin' behaviour of the peptide using UV-resonance Raman spectroscopy, providing more support for the retention of structural elements in the high-temperature unfolded peptide.

The research on Trp cage in a solvent-free environment is notable. The gas phase is an interesting alternative environment for the study of biomolecules; the ease with which an ion can be simulated and the range of gas-phase techniques available render it a good

medium to study (among others) protein and peptide structure in an apolar environment (being much more similar to the hydrophobic phase of a membrane bilayer than water). Still, none of the available techniques is able to determine the gas-phase structure of macromolecules at high resolution. These considerations highlight the need for macromolecular systems that can be addressed both theoretically and experimentally. The obstacles one commonly meets when attempting to elucidate the membrane-bound structures for biopolymers experimentally further intensify this need.

Mass spectrometry of Trp cage ions has been employed to establish links between the miniprotein's conformational preferences in solution and gas-phase data. Lin *et al.*[17] have compared ESI-QToF-MS spectra of Trp cage and a number of variants obtained using different spray solutions (by controlling pH or adding TFE co-solvent) and have shown that charge-state distributions indicate the population of folded peptides, less well-folding peptide constructs and denaturing solution conditions giving rise to a greater population of $[M+3H]^{3+}$ ions. The authors have also used HDX in solution to interrogate the structure of the peptide.

Adams *et al.*[18] have used electron capture dissociation (ECD) to probe the presence of higher-order structure in the vapour-phase Trp cage sequence. The presence of different fragment ions was used to provide an estimate for the location of protons along the peptide sequence; the fragmentation pattern could even distinguish between the all-L peptide and the D-Tyr3 variant. In a more recent study, Patriksson *et al.*[14, 15, 19] have used a replica exchange method to sample the conformational space of the peptide *in vacuo* and assessed the accord between predicted and experimental N-C_α bond cleavage frequencies; the study examined the all-L peptide and three stereoisomers (d-Tyr3, d-Gln5 and d-Leu7). Although a 'perfect' agreement between theoretical and experimental data could not be reached, much of the discord was caused by the complexity of the ECD mechanism and a lack of a single model that can robustly predict ECD mass spectra from an ensemble of simulated structures. However, the authors postulated an alternative ECD mechanism whereby an electron is captured by a backbone amide nitrogen (in spatial proximity to the protonated residue) hydrogen-bonded to a backbone

carbonyl oxygen more remote in sequence, thus stressing the importance of neutral hydrogen bonding in the ECD mechanism[15].

The potential of Trp cage in fluorescence studies has not been acknowledged by condensed-phase experimentalists alone. Iavarone and Parks[20] have probed the vapour-phase conformational transitions of the peptide induced by columbic repulsion and thermal fluctuations using a fluorescence quenching method. The Trp cage was modified by the addition of a lysine residue at the C-terminus to which the BODIPY-TMR fluorophore was attached via a flexible linker. In the solution-phase Trp cage fold the central W6 is "caged" by hydrophobic residues (mostly proline residues); once this core is disrupted *in vacuo* **and** collisions between W6 and the BODIPY-TMR dye become more frequent, the dye fluorescence is quenched. Fluorescence quenching data for the optimised tc5b sequence and the sequence lacking the N9D mutation (Table 3.1) have been reported[14, 20]. These experiments have revealed several aspects of the gas-phase structure of Trp cage. Fluorescence emission was recorded for both peptides at low temperatures, suggesting the presence of a compact gas-phase structure even for the poorly folded D9N peptide. Peptide fold destabilisation as a result of columbic repulsion was demonstrated: for the 2+ charge state the onset of fluorescence quenching took place around 420K, whereas for the 3+ charge state at 360K. It was also found that the optimised sequence was more stable than the D9N construct, since both charge states of the latter began to unfold at lower temperatures (as inferred by a drop in FRET efficiency). Whether this trend was due to retention of the solution-phase fold in a solvent-free environment during experimental timescales or to the presence of an aspartate moiety at position 9 resulting in a more stable gas-phase fold remains unclear.

Trp cage has also been employed as a model for investigating protein solvation at a molecular level. MD simulations of partially solvated Trp cage performed by Patriksson *et al.*[21] have shown that distribution of water molecules around peptide ions is highly anisotropic, due to the existence of surface-exposed hydrophobic regions. These observations are in good agreement with NMR studies of Trp cage in a series of organic co-solvents, indicating co-solvent molecules preferentially aggregating around the peptide in solution[22].

In the current study we aim to extend the investigation of the structure of the Trp cage miniprotein in a solvent-free environment using a combination of molecular mechanics simulations and ion mobility measurements.

The terms “wild-type”, “native”, “original” in the context of Trp cage, will refer to the optimised synthetic sequence tc5b (as labelled by Neidigh et al.[23]) and not to the fragment of the naturally-occurring sequence of gila monster extendin-4; similarly, the characterisation “mutant” will be designated to variants deviating from the tc5b sequence.

3.3 Mass Spectrum of Trp cage

Since gas-phase studies on Trp cage form the core of the entire thesis, the entire project can be thought of as a detailed analysis of the mass spectrum of Trp cage (and its variants).

At near-neutral pH (6.3<pH<6.9) the nano-ESI mass spectrum of the optimised Trp cage peptide (Figure 3.3) displays the following features:

- *Charge-state distribution* The 2+ charge state dominates, the 3+ charge state also being always detected. A low-intensity 4+ peak seldom detected, yet due to the low abundance and sparse, sporadic appearance of 4+ ions their origin could not be elucidated.
- *In-source oligomer formation* Oligomeric aggregates (from dimers to pentamers) are significantly populated in MoQToF mass spectra, especially when the first transfer region is kept at high pressure. In addition to states with non-integer charges per monomer, aggregates also give rise to an $[nM+nH]^{n+}$ (apparent 1+) charge state.
- *Adduct formation* Sodium, potassium and ammonia/water adducts are readily detected and assist, along with (seldom resolvable in practice) isotope distributions, charge- and oligomerisation-state assignment.
- *Fragmentation* Low-abundance CID-type fragments were occasionally detected. Water/ammonia loss (mass difference of 18) was by far the most common rearrangement.

The pre-eminence of the 2+ charge state follows intuitively from the fact that two basic residues are present (K8 and R16). A third proton may be located at the N-terminus, at any backbone carbonyl or amine group (especially favourable for proline residues), and on any of the polar side chains, with N1, Y3, Q5, W6 being particularly attractive contenders. Q5 in particular has been favoured by Adams *et al.*[24] and Patriksson *et al.*[25]. In addition to the overall ionic populations, the magnitude of the $[M+2H+X]^{3+}/[M+3H]^{3+}$ and $[M+H+X]^{2+}/[M+2H]^{2+}$ ratios, the former being significantly higher, show that a 3+ charge-state is less favourable at least when Trp cage is sprayed from aqueous solution at near-neutral pH.

Although persistent, aggregate formation appeared to correlate with concentration and the oligomer distribution did not change significantly over several days: these observations discourage the interpretation of there being a concerted aggregation process occurring in solution, such as amyloid formation. Instead, observed aggregates must be thought of as by-products of the nESI process, which are intensified by the modifications imparted on the source and transfer regions of the MoQToF (which encourage preservation of non-covalent interactions, as discussed in section 2.3). The presence of a significantly populated $[2M+3H]^{3+}$ state suggests that dimers (and possibly hexamers) may contribute partially to the apparent 2+ charge state; this hypothesis was proven by subsequent IMMS analysis. Such an event is of major consequence to IMMS data as a more involved method of analysis is needed to account for the presence of multimers coinciding with monomeric arrival time distributions.

Adduct formation is ubiquitous; yet, contrary to aggregation, the intensity of sodium and potassium adducts increases slowly with time over several days, suggesting that contaminating salt originates in the glassware and plasticware in which samples have been stored.

The ability to differentiate between different species with such accuracy is a trait unique to mass spectrometry, attributing this technique's main advantage over other widespread biophysical approaches. Understanding gas-phase configurations of peptides defines the main impetus of studying Trp cage in such an "artificial" environment; thus in the

sections that follow the analysis will concentrate mostly on 3+ and 2+ monomeric species.

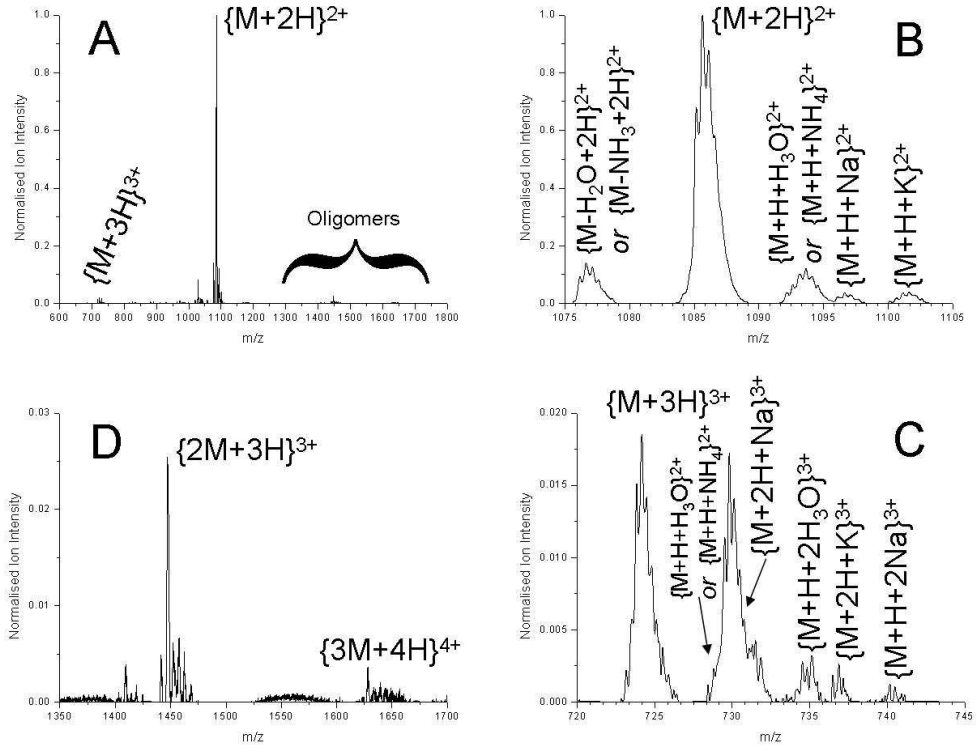


Figure 3.3
Mass spectrum of Trp cage tc5b (WT). A) Full spectrum showing $[M+2H]^{2+}$ and $[M+3H]^{3+}$ ions, as well as oligomeric aggregates formed in-source. B) Doubly-charged ions; water/ammonia, sodium and potassium adducts as well as a dehydrated species are clearly distinguishable. C) Triply-charged ions, showing water/ammonia and sodium/potassium adducts. D) Putative dimeric and trimeric ion aggregates.

3.4 Mobility of Trp Cage Ions in Helium

IMMS experiments on Trp cage confirm the conclusions drawn from the appearance of the mass spectrum and enable the determination of drift velocities orientationally-averaged momentum transfer cross-sections (from expression 1.23) which are subsequently compared to theoretical results. Cross-sections of Trp cage ions are listed in Table 3.2. Analysis of IMMS experiments revealed the following trends:

- *Collapse into a more compact gas-phase configuration* With the NMR structure[23] of tc5b yielding a collision cross section of 420Å^2 (calculated with *mobcal* using the trajectory method[26]) it is evident from the experimentally determined momentum-transfer cross sections that all monomeric species collapse into compact geometries upon removal of solvent.
- *Coulombic repulsion drives gas-phase unfolding* As shown in Table 3.2, $[M+3H]^{3+}$ ions are characterised by larger collision cross sections compared to $[M+2H]^{2+}$ ions, underlining the importance of coulombic effects in determining the conformation of unsolvated peptide ions.
- *Possibility of multiple co-existent conformations* Fitting experimental data to equation 1.30 consistently yields apparent diffusion coefficients that are significantly larger than those predicted from the Nernst-Einstein-Townsend relation $D = K\frac{k_B T}{q}$ (in other words, ATDs are much broader than expected). Although several factors (such as the width of the injected ion packet, diffusion outside the drift region, end effects and space-charge effects) may well contribute to peak broadening, the presence of multiple conformers may also be responsible for such a phenomenon.
- *Presence of $[2M+4H]^{4+}$ dimer (with the possibility of other $[2nM+2nH]^{2n+}$ oligomers)* In addition to being broader, ATDs of $[2nM+2nH]^{2n+}$ ions separate in two peaks at low drift fields. An example of this is shown in Figure 3.4a,b. The existence of multimeric species in the spectra provides evidence for the early-arriving species being a dimer. The compact nature of the monomers also corroborate this hypothesis (interpreting the early-arriving species as a monomer

would attribute to it a momentum transfer cross section that is too low to be physically reasonable).

- *Effect of sodium* As charge solvation (the stabilisation of a charged moiety by maximising the interactions of the charge with moieties with partial charges of the opposite sign) appears to influence the ionised miniprotein's structure in the absence of solvent, metal ions, by virtue of not being covalently bound to the protein chain, can form the core around which conformational rearrangements occur in order to stabilise its high charge density. Even "non-specific" metal ion adducts can be of thought as order-inducing, yielding compact configurations. However, Trp cage ions appear to be less sensitive to such effects at 300K. $[M+H+Na]^{2+}$ cross sections are similar to those of $[M+2H]^{2+}$. $[M+2H+Na]^{3+}$ ions are found to be more extended compared to their non-halogenated counterparts – therefore in this case the sodium adduct accentuates Coulombic repulsion-mediated unfolding effects.

Table 3.2
Collision cross sections of Trp cage ions at ambient temperature (≈300K)

Ion	*m/z*	*Collision Cross Section / Å²*
$[M+2H]^{2+}$	**1086**	**366.0 ± 20.2**
$[M+3H]^{3+}$	**724.1**	**391.7 ± 11.3**
$[M+H+Na]^{2+}$	**1097**	**361.9 ± 29.6**
$[M+2H+Na]^{3+}$	**731.5**	**408.1 ± 10.6**
$[2M+3H]^{3+}$	**1447**	**599.2 ± 4.5**
$[2M+4H]^{4+}$	**1086**	**590.3 ± 134.0**

Having established these relationships by ion mobility experiments of Trp cage, we now shall proceed to make structural assignments of the calculated cross-sections by means of molecular modelling strategies, starting with studies employing simulated annealing, an energy optimisation technique.

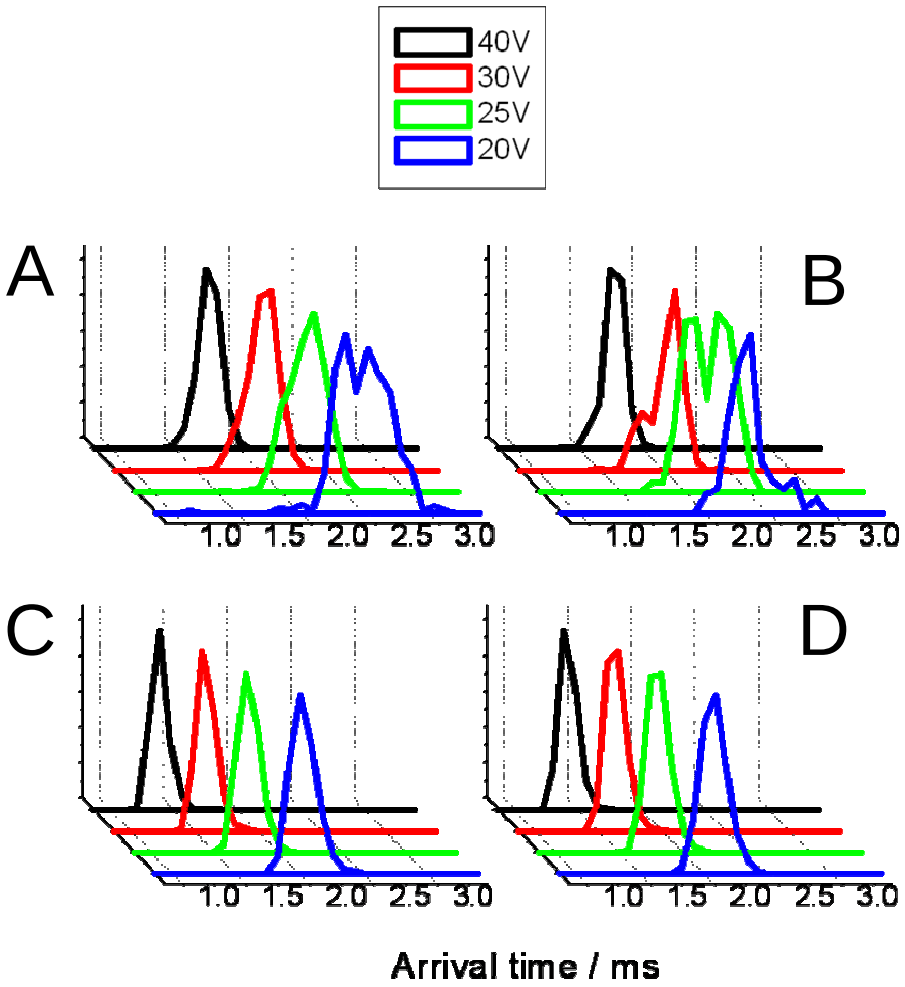


Figure 3.4
Arrival time distributions for Trp cage ions. A) $[M+2H]^{2+}$; B) $[M+H+Na]^{2+}$; C) $[M+3H]^{3+}$; D) $[M+2H+Na]^{3+}$

3.5 Gas-phase Structure of Trp Cage Ions as Predicted by a Simulated Annealing Algorithm

Attempts to simulate isolated ions containing several ionisable groups using atomistic models must face the issue of locating the charge. As stated above, Trp cage contains two basic residues, K8 and R16, these two being offered as straightforward choices for the most dominant charged sites of the 2+ cation. A third charge can be located at the N-terminus, yet there are alternatives; most notably, Adams *et al.*[27] have provided support for Q5 being charged (in fact their preferred assignment, based on ECD data, place the charges at the N-terminus, Q5 and R16, leaving K8 neutral). Trp cage also contains an acidic side-chain, that of D9, which in solution allows the formation of a salt bridge involving K8 and R16, which has been shown to be of great importance for stabilising the "native" fold[23, 28, 29]. Although it is known that zwitterions are unstable *in vacuo* (to illustrate this fact, the highly basic guanidino group of arginine in $[arginine:Na]^{1+}$ complex is neutral[30]) zwitterions can potentially be stabilised by molecular conformation, as suggested for bradykinin $[M+H]^+$ ions[31]. A systematic survey of charge locations should envisage all possibilities. Yet such an approach quickly becomes impractical due to the large number of possible charge locations and the small differences these would produce, which are unlikely to be distinguishable experimentally. A more realistic strategy would allow protonation sites to change with simulation time; yet such an approach is beyond the reach of the relatively simple models employed herein. Thus only a few select cases which were deemed important were considered. Now, Trp cage, like bradykinin, is proline-rich and as the salt-bridge plays an important role in the solution structure, it was considered an attractive case for the 2+ ion. R16, K8, the N-terminus and Q5 were the other sites considered for charge assignment; again, only a subset of combinations was modelled, which is listed in Table 3.3.

Table 3.3
Charge assignment of modelled peptides (free termini).
Cationic residues are shown in blue and anionic residues in red.

Label	*Net Charge*	*Proton location*
WT2a	**2+**	**NLYIQWLKDGGPSSGRPPPS**
WT2b	**2+**	**NLYIQWLKDGGPSSGRPPPS**
WT2z	**2+**	**NLYIQWLKDGGPSSGRPPPS**
WT2z-b	**2+**	**NLYIQWLKDGGPSSGRPPPS**
WT3a	**3+**	**NLYIQWLKDGGPSSGRPPPS**
WT3b	**3+**	**NLYIQWLKDGGPSSGRPPPS**
WT3c	**3+**	**NLYIQWLKDGGPSSGRPPPS**

All constructs were subjected to 300 iterations of the simulated annealing procedure described in Chapter 2 starting from a completely extended structure; parameters for the simulated annealing runs are displayed in Table 3.4.

Table 3.4
Simulated annealing parameters for Trp cage

High Temperature	**800K**
Minimum Temperature	**0K**
Length of high T MD	**30ps**
Length of cooling MD	**20ps**
Length of MD step	**1fs**
Number of cooling steps	**20**
Number of iterations	**300**

Collision integrals were estimated for all structures using *mobcal*, as already described. Yet apart from the cross-section, it is also useful to define another structural parameter which may enable a concise description of structural characteristics. Therefore, we may define the per residue helical propensity as

$$A_{Hel} = e^{-(\ln 2)\left[\frac{\left((\varphi-\varphi_0)^2+(\psi-\psi_0)^2\right)}{\sigma^2}\right]^{\alpha}} \qquad \{3.1\}$$

where φ and ψ represent backbone dihedral angles (the subscript "0" indicating the value corresponding to the "ideal" structure) and σ the value at which the function equals $1/2$. Any secondary structure can be represented in this fashion by choosing appropriate φ_0 and ψ_0 (for an α-helix they are set to φ_0 = -60° and ψ_0 = -45°). Such a modified Gaussian obtains high values (close to 1) around (φ_0,ψ_0) and drops relatively steeply around φ_0,$\psi_0\pm\sigma$, the parameter α controlling the steepness of the transition. For a measure of α-helicity the parameters α and σ have been set to 4 and 20° respectively as these values perform well when empirically tested on model helical peptides. Likewise, the similarity to the backbone dihedrals of a reference structure is defined using a 2-dimensional Gaussian (setting α = 1 above) and choosing φ_0 and ψ_0 to the values of the reference geometry. In this fashion the "similarity to the NMR structure" was estimated, using the first conformation of the published tc5b NMR structure (pdb code 1L2Y) and setting σ = 15°. Performing these simple calculations provides a way to represent qualitatively (and yet with proper normalisation) and summarise the local structural preferences of the peptide backbone. Finally it must be noted that data seemed significantly more consistent when statistical averages of the structure per residue over the entire simulated ensemble were performed, without any energy rescaling. Since only a relatively small number of low-energy structures contributes significantly to the average after rescaling, such behaviour can be attributed to the convergence characteristics of the algorithm, which have been expanded upon in section 2.2.2.4.

Being thus mindful of the approach's restraints, one can still make several useful observations from these results. Trends of average local structure (similarity to NMR backbone structure and helical propensity, as already defined) and their dependence on proton location can be viewed in Figure 3.5.

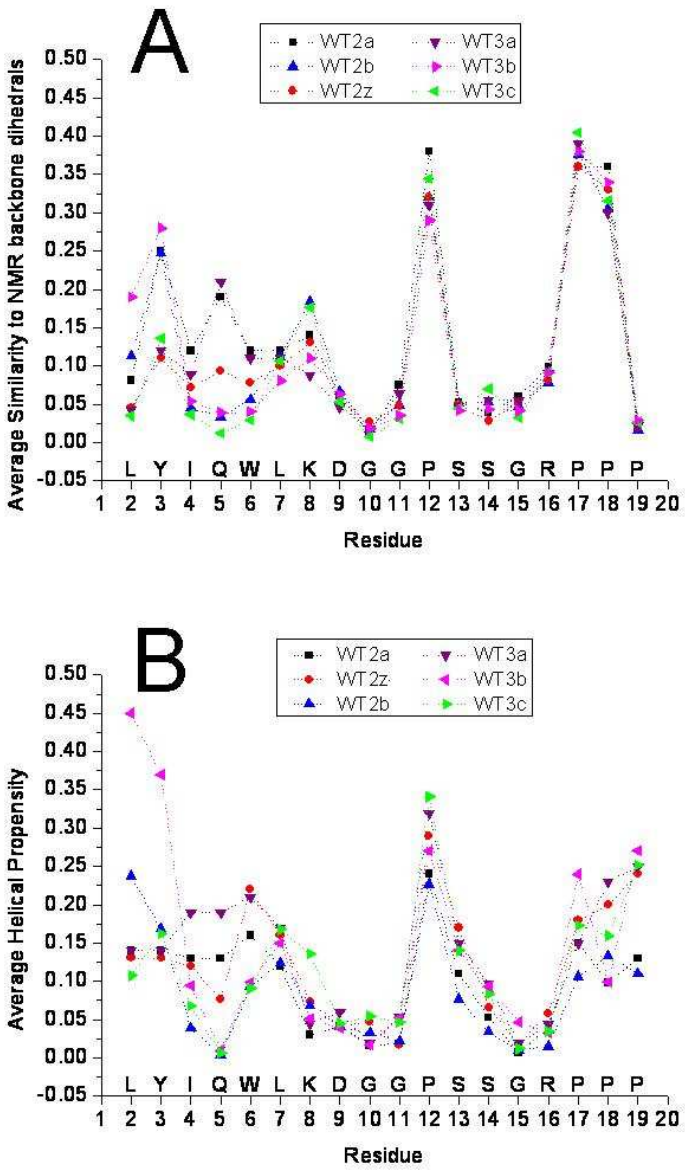


Figure 3.5 Structural characteristics of simulated peptide backbones. A) Similarity to NMR structure B) Helical character. These quantities were calculated using equation {3.1} which normalises values to 1 (the latter value being attributed to the reference geometry). See text for discussion.

Residues 9-17 behave similarly in all protonation states simulated. Conformationally rigid proline residues form either extended ribbon-type structures or facilitate the formation of β-turns (it must be noted that residues i+1 of type I turns adopt gauche+

geometry which gives rise to helix-like (-60, -30) dihedrals and will therefore appear as having a helical character according to equation {3.1}; this fact explains the presence of some apparent helical propensity at residues 17-19). Pro12 in particular seems to assume such a gauche+ configuration, constituting the core of a turn close to the middle of the sequence. A β-turn in the middle of the peptide chain facilitates a collapsed U-like structure on the basis of which further rearrangements may occur. Glycine residues can occupy a large range of conformations and do not settle in either a helical one or the one present in the NMR structure in simulations of this level of convergence. Yet it is in the remaining sequence – especially the N-terminal 1-8 segment (forming an α-helix in the NMR structure) – in which most pronounced differences between charge assignments can be seen:

- Residues 2-5 of the zwitterionic WT2z peptide are less helical than the WT2 form; however, the inverse is true for residues 6-8.
- The presence of a charge at the N-terminus destabilises helix formation; yet a salt bridge between K8 D9 and R16 brings the two sides of the peptide in close proximity which in turn induces compact helical turns to form locally on either side of the peptide.
- Eliminating the charge of D9 in WT3a increases the helical propensity upstream K8 indicating solvation of the lysine's charge in proximal backbone carbonyl groups.
- Locating the proton on Q5 instead of the N-terminus in WT3b promotes helix formation upstream but reduces the helical character of residues 4-7, a trait attributable to the repulsion between the charged Q5 and K8 when located on the same side of an α-helix.
- However, when the N-terminus and Q5 are charged, K8 being neutral, the helix and consequently native-like character of the peptide is diminished.
- As far as the similarity of simulated structures' backbone configuration to the solution NMR fold is concerned, WT2 retains such characteristics at a greater extent than WT2z at residues 1-7 and WT3a is more similar to the NMR structure than WT3b at residues 4-8.

- Both WT3 constructs differ mostly from WT2 around the additional charge.
- Simulations where the Q5 side-chain bears a charge generally disrupts helicity in the vicinity of Q5; yet in the case of WT3b the helical character *upstream* of Q5 is greatly enhanced. This effect can be attributed to the ability of a positive charge located at the C-terminal end of a peptide sequence to enhance helix formation (charge localisation at or near the N-terminus yielding the opposite result) as appropriate helix-initiating hydrogen bonds can thus be established[32].

Collision cross sections of the simulated geometries (calculated with *mobcal*) underline the dominance of electrostatic interactions on gas-phase conformation (Figure 3.6).

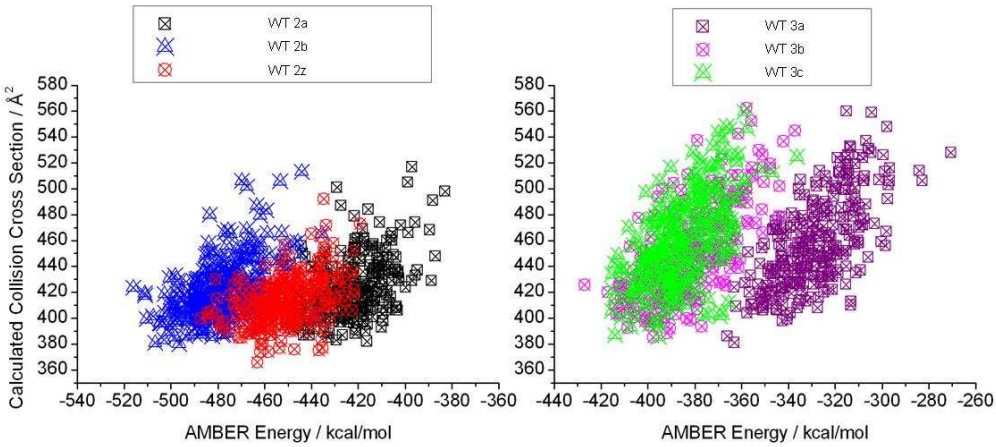


Figure 3.6
Collision cross section, calculated with *mobcal* using the trajectory method, vs. force-field energy for different charge assignments of $[M+2H]^{2+}$ and $[M+3H]^{3+}$ WT Trp cage. Zwitterionic systems (red circles) give rise to smaller and more tightly distributed collision cross-sections. 3+ ions produce more diffuse collision cross-section distributions than 2+ species, indicative of the effect of coulombic repulsion. Positioning the charge on Q5 instead of K8 does not affect the cross section distributions significantly.

As expected, WT2 is characterised by larger cross sections than zwitterionic WT2z, preservation of a salt bridge *in vacuo* leading to the formation of more compact structures. The zwitterion potential energy is also lower compared to WT2 (although AMBER energies contain contributions from residues with dissimilar parameters and, being empirical force-field energies, do not account quantitatively for electronic

structure energy contributions). The distribution of numerically calculated momentum transfer integrals is also narrower for the zwitterionic $[M+2H]^{2+}$ peptide compared to WT2. $[M+3H]^{3+}$ peptide constructs routinely possess larger cross-sections and also give rise to significantly broader cross-section distributions than 2+ charge states; however it must be noted that the low-lying energy minima sampled by simulated annealing for WT3a yield a collision cross section somewhat lower than that of WT2 (Table 3.2). On a fundamental level it can be said that strong electrostatic attractions will give rise to more compact structures whereas repulsion between like charges will cause the peptide to unfold. Simulated cross-sections obtained with a charged Q5 are similar to these resulting when Q5 is kept neutral; when an exponential energy weighting factor is applied, (equation {1.18}) the resulting cross sections for peptides with charged Q5 are somewhat larger than averages calculated for constructs where Q5 is not charged.

Comparing calculated to experimental cross sections it is apparent that experimental values are invariably lower than those predicted from simulation, hindering precise charge assignment at this point. The reason for this observation is not known. Convergence problems of the optimisation algorithm may be one reason for obtaining such results. Injection energy effects can be discounted: measurements collected at higher injection energies yielded higher cross sections, indicating peptide unfolding. Misassignment due to aggregation is also not responsible for this effect since fitting of the ATDs with equation {1.30} to account for numerous non-interconverting species does not resolve the discrepancy. This issue must undergo further investigation, and the possibility that force field parameters are to blame must be considered. Electrostatics play a dominant role in determining gas-phase conformation, and if the charge on a given atom/functional group is underestimated this will result in ‘looser’ conformations, which would be consistent with the effect seen. Nevertheless, all qualitative trends reported here agree with experiment and can be explained by the existing corpus of knowledge on solvent-free peptide structure (reviewed in section 1.5).

When single experimental measurements are in quantitative agreement with simulation results intriguing observations can be made. Measured ATDs are much broader than expected which can be demonstrated by relating ionic mobility to diffusion coefficient

via the Nernst-Einstein-Townsend relation (as discussed in Chapter 1 and Appendix B). By adding ATDs calculated with equation {1.30} it can be seen that the observed broadening of a given ATD can be assigned to the presence of multiple unresolved conformations. This approach is equivalent to assuming that the sampled energy minima adequately represent the observed ensemble and that the energy barriers separating them are too large for interconversion to take place at the low millisecond timescale. Figure 3.7 shows an example of this for $[M+3H]^{3+}$ where M corresponds to Trp cage.

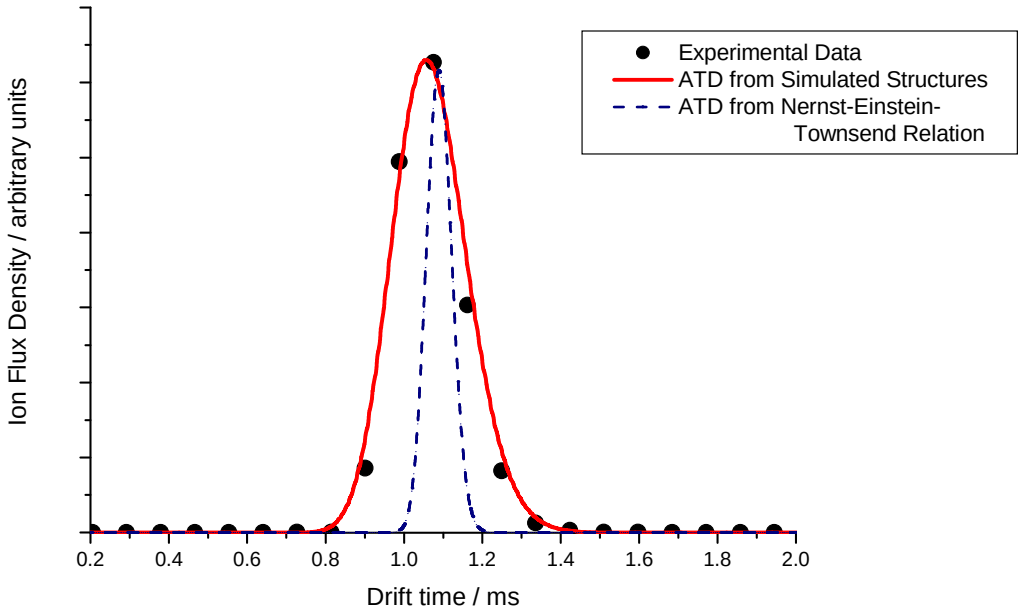


Figure 3.7
Simulation of arrival time distribution for WT $[M+3H]^{3+}$ (this particular dataset was chosen as it represents one case when experimental and simulated cross sections happened to agree). Black dots: experimental data. Dashed line: ATD calculated from equation {1.30} using a mobility identical to that determined experimentally and estimating the diffusion coefficient by the Nernst-Einstein-Townsend relation (such a treatment assumes drift of a single species). Red line: weighted sum of ATDs estimated with the same equation but using theoretical cross sections calculated for geometries generated by simulated annealing. The exponential weight factor is described in section 1.4. This treatment assumes a mixture of non-interconverting conformations, each with population determined by the aforementioned weight factor. The widening of the ADT observed can therefore be attributed, among other factors, to the presence of numerous configurations.

Visual inspection of simulated annealing low energy structures can further illustrate the trends described above. Snapshots of simulated structures are shown in Figure 3.8. All constructs adopt geometries containing a turn close to the middle of the sequence with

flanking segments being ornamented with further turns. Organised secondary structure is seen frequently but not ubiquitously; 3.8b and 3.8d present particularly interesting cases since their backbone geometry shows some similarity to the NMR structure. The central U-shaped topology caused by the turn around Pro12 emphasises the significance of self-solvation in stabilising the configuration of these peptides; however, as seen in Figure 3.8 and by inference (Figure 3.5), the relative orientation of the N-terminal, central and C-terminal regions, essentially controlled by the turns at residues 9-11 and 14-16 are not reproduced. Gly10 and Gly15 possess on average nearly no similarity to the NMR structure. As these turns determine the relative positioning of the different secondary structural elements and, consequently, the peptide's tertiary structure, we can see that *in vacuo*, tertiary elements, brought together and stabilised in solution by van der Waals interactions between hydrophobic amino acid side-chains, can not form spontaneously in the absence of solvent.

As results from simulated annealing are compromised by the convergence properties of the algorithm, a longer simulation, thereby allowing better convergence, was performed. Due to the computational expense of this approach only a single run was performed, for the WT2 construct, and only for 10 iterations. Simulated annealing parameters for this run were chosen on the basis of the convergence tests discussed in Chapter 2 and are presented in Table 3.5.

Snapshots of four lowest-energy structures resulting from this run are shown in Figure 3.9 whereas energies and cross-sections of all resultant geometries are displayed in Figure 3.10. The occupation of lower-lying energy minima by this simulation is immediately apparent; notably, these lower-energy structures do *not* give rise to calculated cross sections that are significantly lower than those of low-energy minima generated by the more hasty simulation. Low-energy structures from this latest run, snapshots of which can be seen in Figure 3.9, share considerably more characteristics than discerned from the earlier simulation. First of all, much longer heating and cooling times allow for more well-defined elements to occur. The peptide can still be divided into the three regions which also define its solution fold: an N-terminal helix, a 3_{10}-helical turn and a more extended segment containing the three proline residues, all three

elements being separated by turns. Helical elements do form *in vacuo* but their solution-like relative orientation, i.e. tertiary structure, is never visited from the algorithm, a conclusion that has already been reached on the basis of the results presented earlier in this section. Thus it can be said that simulated annealing allows for the general characteristics of the solvent-free structures of the studied systems to be outlined. Secondary structure certainly seems more robust than tertiary structure *in vacuo*. The following questions can now be posed:

1. To what extent will the solution structure of the Trp cage miniprotein change once relieved from solvent?
2. Does charge assignment influence this trans-conformational event?

These issues are addressed via molecular dynamics (MD) simulations.

Table 3.5
Parameters for long simulated annealing runs of Trp cage

High Temperature	**800K**
Minimum Temperature	**0K**
Length of high T MD	**500ps**
Length of cooling MD	**55000ps**
Length of MD step	**1fs**
Number of cooling steps	**20**
Number of iterations	**10**

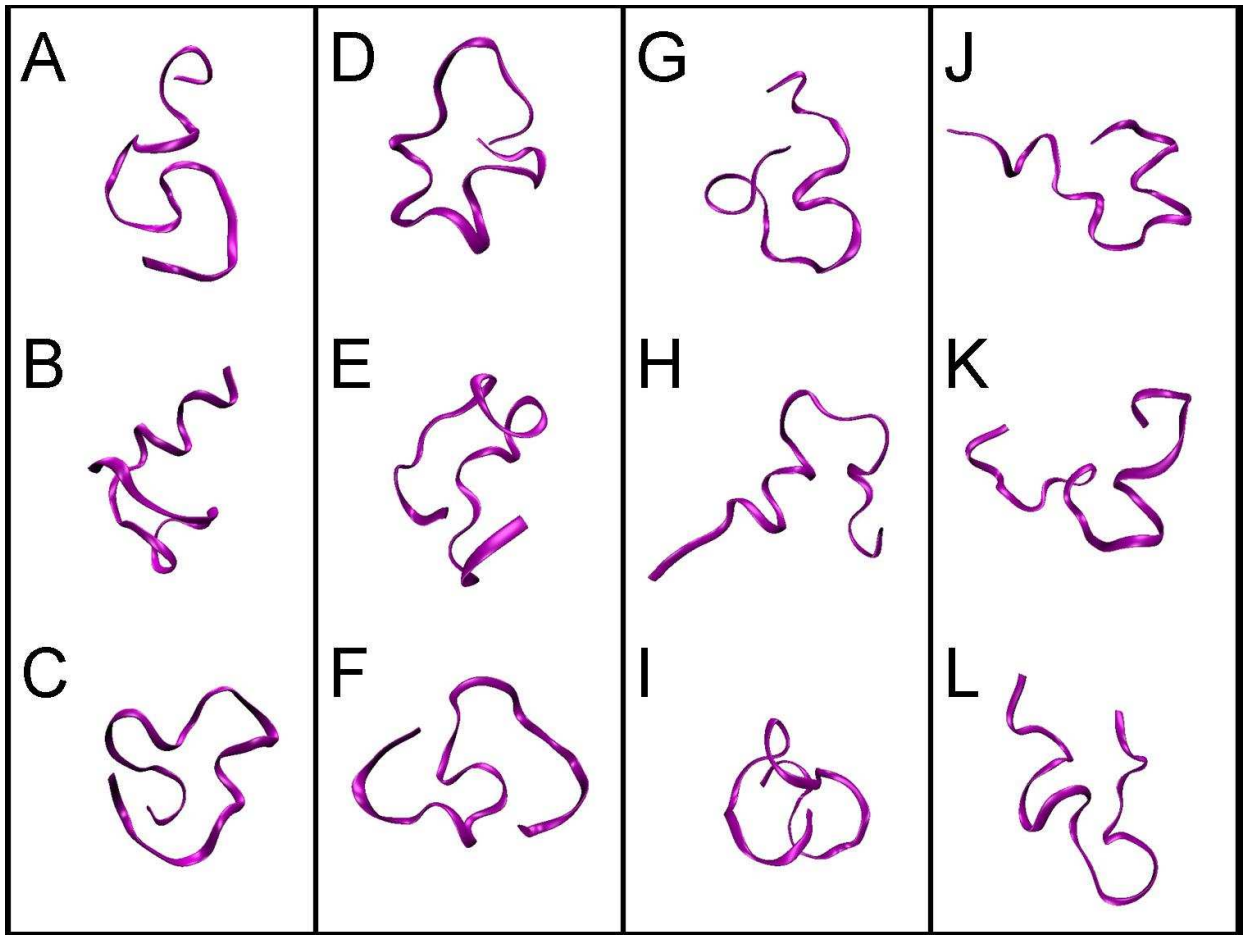


Figure 3.8
Low-energy structures of Trp cage generated by simulated annealing.
A-C) WT2a D-F) WT2z G-I) WT3a J-L) WT3b

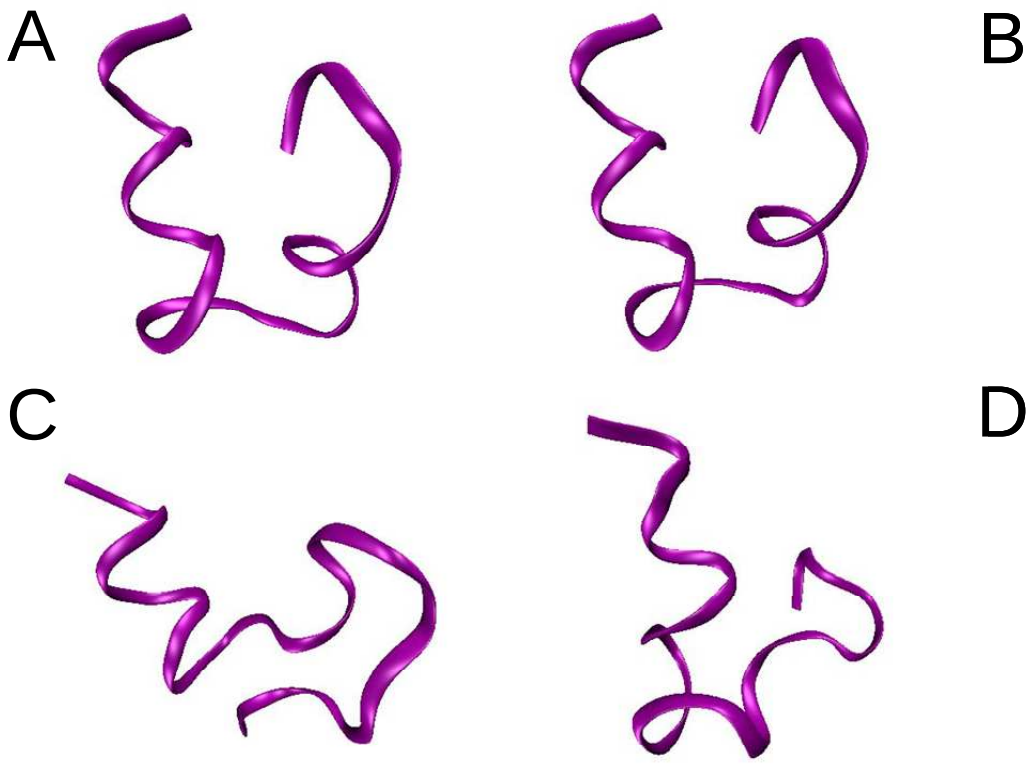


Figure 3.9
Four lowest-energy configurations generated by simulated annealing of the WT2a construct using long relaxation times. Notably the helical nature of the N-terminal segment is present throughout. The two lowest-energy structures (A&B) display great similarities and also reproduce the U-shaped topology characteristic of the NMR structure.

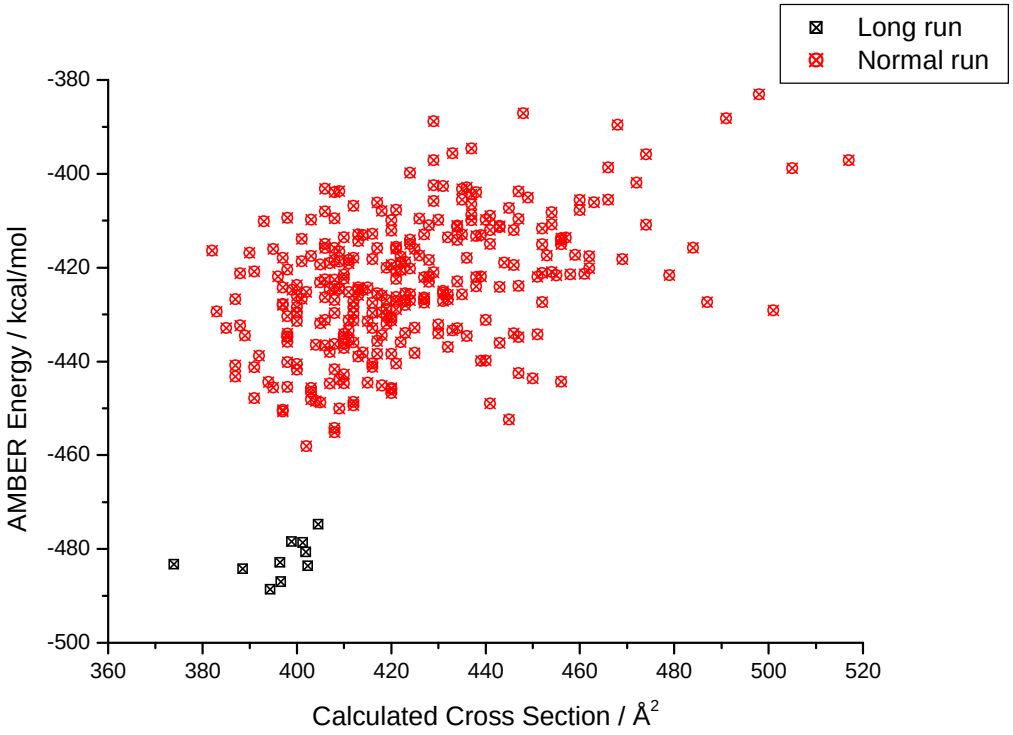


Figure 3.10
AMBER energy vs. calculated cross section scatter plots for geometries of WT2a generated with very long annealing times (Table 3.5) allowing for convergence (black squares) compared to geometries generated from parameters displayed in Table 3.4 (red circles). The two calculated cross-section distributions are significantly different but, notably, they do overlap.

3.6 Stability of the Trp cage fold *in vacuo* as probed my MD simulations

Structural optimisation by simulated annealing in the absence of solvent does point out essential features stabilising gas-phase conformations. Yet SA cannot, by its very nature, directly characterise the stability of a 'native' solution fold *in vacuo*. On the other hand, MD simulations can provide detailed pathways of the conformational transitions of a given initial configuration and are, consequently, ideally suited for the aforementioned task. Although MD simulations may not converge to the global minimum, they can be used to sample conformations neighbouring the starting structure in conformation space. Thus the assumption that whatever transitions occur within the few nanoseconds of simulation time will most certainly take place during the milliseconds of experimental measurement, given that the initial geometry is a realistic one, becomes justified. MD runs were carried out for the charge assignments listed in Table 3.2, using the first geometry of pdb file 1L2Y[23] for the initial structure. The initial geometry was parameterised, energy-minimised and submitted to the gradual heating protocol, relaxation MD and production MD at a range of temperatures (10ns of MD using a 0.5fs time-step at simulation temperatures in the 300-600K range, heating between the runs, yielding a total simulation time of 70ns, i.e. $1.4 \cdot 10^8$ MD iterations). The simulation protocol is described in greater detail in Chapter 2. We shall proceed with summarising the MD simulations at 300K. Both experimental and computational data at higher temperatures are presented in Chapter 4 so a more detailed analysis of the MD simulations can be postponed until then. For the time being we will limit our analysis to a comparison of the simulation trajectory at 300K with the Trp cage fold's backbone configuration, their theoretical collision cross-sections (calculated with *sigma* every 10ps, as described in section 1.5.5) and the distance between the peptide termini – parameters which suffice for illustrating the most pronounced effects brought about by different charge assignments on the 'wild-type' fold at the said temperature. These data can be seen in Figure 3.11.

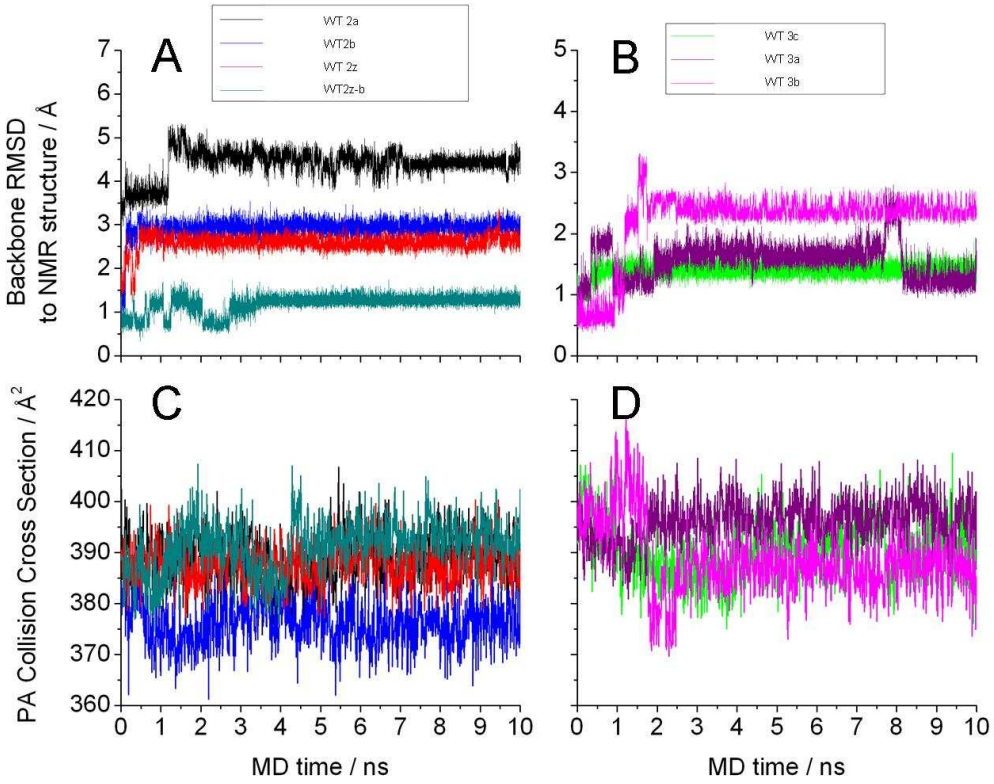


Figure 3.11
Heavy backbone atom RMSD from the NMR structure and calculated cross sections for molecular dynamics simulations of Trp cage (WT, free termini) A-B) Backbone RMSDs. C-D) Collision cross sections calculated with *sigma* every 10ps. A,C) $[M+2H]^{2+}$ ions. B,D) $[M+3H]^{3+}$ ions.

MD simulations at 300K indicate that charge assignment will influence significantly fold stabilisation in the gas phase. WT2b preserves native-like characteristics more efficiently than WT2a, which has undergone significant conformational rearrangements during the temperature equilibration stages and continues doing so at 300K. Yet, the Trp cage fold survives best in the case of the two zwitterionic constructs. In particular, zwitterionic termini (WT2z-b) effect the conservation of the solution fold most efficiently. Interestingly, simulated 3+ peptides are not less effective at conserving the solution fold than 2+ ions. Similarity to the 'native' backbone geometry is manifest in all three 3+ constructs, in particular WT3a and WT3c (the former displaying somewhat

greater fluctuations). Within 2ns of the simulation, WT3b undergoes a conformational change, reflecting the repulsive forces acting between the three proximally located charges. The simulation data support the hypothesis that strong electrostatics hinder large-scale conformational rearrangements, as self-solvation constitutes the strongest driving force for such events. In particular, more potent electrostatic interactions between the termini (ionic interactions in WT2z-b or charge-dipole interactions in WT3a,c) play a major role in stabilising the Trp cage fold *in vacuo*.

Notably, conformational changes do not give rise to more extended structures; in fact the contrary is true. Structural rearrangements cause the systems to collapse into more compact geometries, characterised by smaller cross sections; the peptide chain *may* only become more extended during a conformational change, but will collapse further as the system equilibrates in a different 'attraction basin'. Such structural metanarratives can be clearly seen in the case of WT2b, WT3b and WT3c. MD simulations agree with the qualitative differences between charge states; 3+ peptides do indeed give rise to higher cross sections than 2+ ions. The solution structure is more extended than the gas-phase structures and removal of solvent can effectually result to a structural collapse. However the rate at which this event happens is dependent on charge localisation. Thus, although zwitterionic constructs yield more compact geometries by simulated annealing, they occupy less extended configurations in these gas phase trajectories due to their property of stabilising the solution fold.

3.7 Factors contributing to the gas-phase conformation of Trp cage

IMMS and molecular modelling studies of Trp cage have succeeded in revealing fundamental features of the miniprotein's solvent-free façade. Experimental evidence for the compact nature of the peptide, unfolding mediated by coulombic repulsion and the organising effect of sodium adducts of the 2+ charge state has been presented. Modelling of Trp cage ions in the gas phase provides support for persistence of helical elements in the absence of solvent on the basis of which one can reasonably suggest that charge location is by far the most important factor in influencing Trp cage conformation

in this minimalistic environment. 3+ ions occupy more extended intermediate-energy structures than 2+ ions; furthermore zwitterionic constructs enable the backbone topology to be greatly conserved during MD simulations performed, whereas constructs lacking such interactions may suffer significant rearrangements at low temperatures. Hydrophobic interactions seem to play a secondary role as simulated annealing structures never exhibit turns present in the solution structure.

As charge localisation appears to be of great significance, three Trp cage variants were synthesised, containing Arg to Lys and Lys to Arg substitutions and subjected to the same analysis. Since an interaction between the two peptide termini was quite prominent in MD runs, especially when one or both are charged, variants with free termini and protecting capping groups (acetylation of the N-terminus and amidation of the C-terminus) were produced for all four constructs as well.

Likewise, as Pro12 plays an important role in the solution structure and, due to the rigid nature of the imino acid, its local structure is reproduced by simulated annealing, a construct containing a d-Pro at position 12 was also investigated. Only the *d*-Pro12 variant with 'free' termini was synthesised.

3.8 Studies on Trp cage Variants with Different Gas-phase Basicity

Gas-phase basicity of amino-acids making up the protein of interest will have direct bearing on charge (de-)localisation along the peptide chain. Although there can be little doubt that nearby residues (proximal in *space* and not merely in sequence) can alter the basicity of an amino acid, basicities of isolated amino acids are a truly indispensable starting point. Gas-phase basicities of individual amino acids as reviewed by Harrison[33] can be seen in Appendix A. That arginine has a greater proton affinity than lysine is a well known fact, not only supported by chemical intuition but also a large corpus of experimental findings[33]. Trp cage construct tc5b (herein termed – however inappropriately – "wild type", WT) contains two strongly basic sites, Lys8 and Arg16. To investigate *whether* basic amino acids with slightly different proton affinities can influence gas-phase structure, all three possibilities of Arg to Lys and Lys to Arg

substitutions were designed and synthesised, with both 'free' and capped termini. The constructs are listed in Table 3.6.

Table 3.6
Trp Cage Constructs containing Arg and Lys Substitutions

Construct	*Sequence*
WT	**NLYIQWL**K**DGGPSSG**R**PPPS**
K8R	**NLYIQWL**R**DGGPSSG**R**PPPS**
R16K	**NLYIQWL**K**DGGPSSG**K**PPPS**
DM	**NLYIQWL**R**DGGPSSG**K**PPPS**

Assuming that gas-phase basicity of the entire peptide chain depends linearly on constituent residues' basicities, the constructs' relative basicities should be K8R> WT= DM> R16K. Consequently, the strength of the K8-D9-R16 salt bridge will vary among the constructs in the same manner, *given that D9 is indeed charged*. Before we ensue with a description of the solvent-free behaviour of these synthetic 'mutants', we shall first consider the effect of the substitutions on the solution structure by circular dichroism (CD) spectroscopy.

3.8.1 Confirmation of the Solution Conformation

CD spectra of all eight constructs were collected in both the far-UV (190-250nm) and near-UV (260-330nm) region (protocol details are described in the previous chapter). Qualitative features of the spectra (Figure 3.12) are indistinguishable to those already reported[14, 17, 27]. Far-UV spectra contain two broad bands, one at 222nm, characteristic of helicity, and one more intense band at 206nm, which contains contributions from both the helical elements and the PPII ribbon of residues 17-20[29]. At the near-UV range (where the characteristic signature is brought about by UV-absorbing side-chains conformationally 'locked' in a chiral environment[34]) peptides absorb between 260 and 305nm. At least three dichroism bands can be discerned: one peaking at 282nm (assigned to Trp6); a broader band with a maximum at 278nm with a 'tail' trailing towards lower wavelengths (attributable to Y3); finally a third negative band at 293nm.

This latest band can only be observed at three constructs containing an arginine residue, being more intense for K8R; therefore it may be associated with the guanidinium group although further research would be necessary to ascertain this claim. CD measurements prove that all constructs contain significant populations of the Trp cage solution fold in 10mM ammonium acetate and near-neutral pH. Moreover, differences between the relative intensity of different bands hint to constructs' relative stability in solution. To facilitate the comparison, Table 3.7 shows the ellipticities of the constructs at two key wavelengths divided by the corresponding ellipticity of WT (free termini). The wavelengths chosen were 222nm (helicity dichroism maximum) and 282nm (Trp dichroism maximum – which can be interpreted as a probe of tertiary structure since Trp6 is located in the heart of the fold).

The order of intensity of the 222nm band is K8R > WT ≈ DM > R16K, irrespective of whether the termini are protected or not. The same order of intensity is observed for the 282nm band for free termini constructs; however when termini are protected the order becomes DM > WT ≈ R16K > K8R. In three cases the order is identical to that of the constructs' basicity suggesting that the strength of the B8-D9-B16 (B here standing for either K or R) influences directly the stability of the N-terminal helix and – in the case of peptides with unprotected termini – the stability of the overall fold. Interestingly, peptides with free termini display a direct synergy between secondary and tertiary structure. The pattern which arises from constructs with capped termini however cannot be interpreted in the absence of further investigation. Yet another difference between protected and unprotected peptides also appears consistently: capped peptides give rise to more intense near- and far-UV spectra (K8R being the only exception). Ionic interactions between the termini when the peptide is found in the zwitterionic form (generally accepted as the most probable form in an aqueous solvent) may disturb the structure locally or even stabilise partially-folded configurations where the termini are still interacting. Moreover, the presence of an acetyl capping group at the N-terminus forms an additional hydrogen bond when the peptide conforms to a helical geometry; in the absence of the protecting group such a helix-stabilising hydrogen bond is provided by the Asn1 side-chain[23] which, however, can only occupy a small number of configurations for this arrangement

to take place. When this interaction does not occur, the Asn1 side-chain can interact with Lys8, an interaction which is particularly tenacious in gas-phase MD simulations of the peptide.

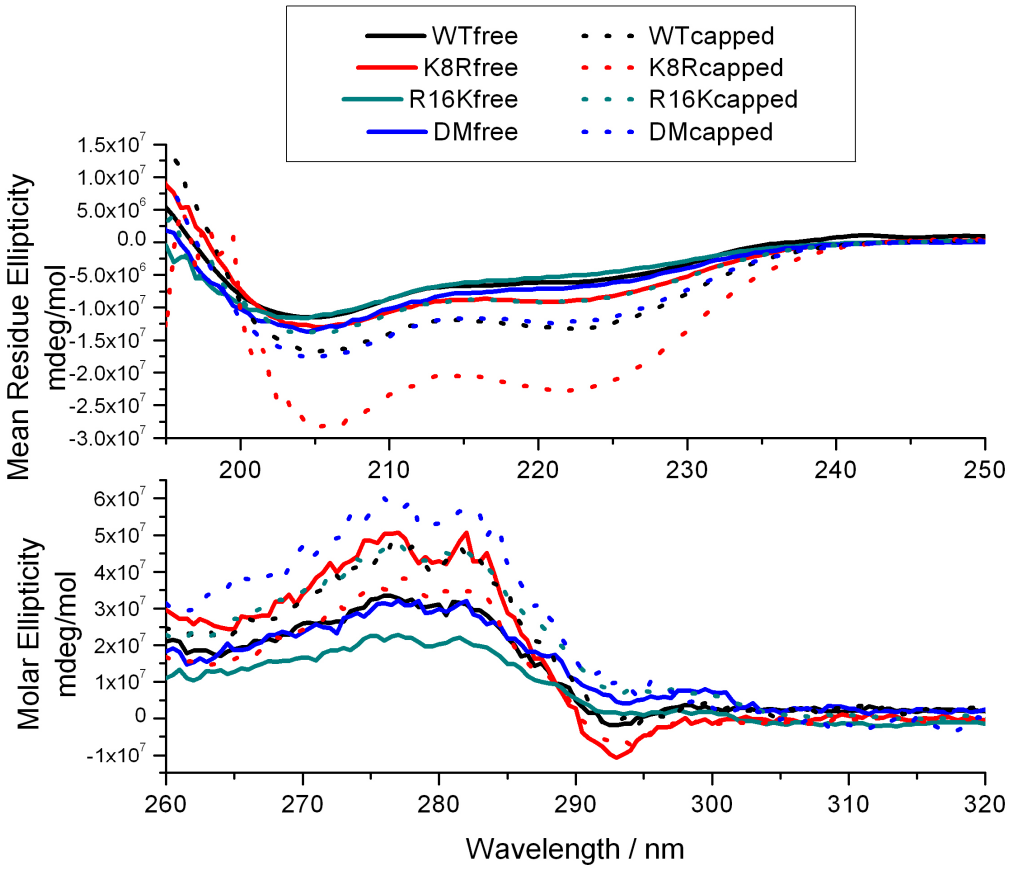


Figure 3.12 Circular Dichroism spectra of Trp cage constructs in 10mM NH_4CH_3COO, pH 6.7, at 298K. Top: far-UV spectra; Bottom: near-UV spectra. Solid lines: peptides with free termini. Dashed lines: peptides with protected termini.

Table 3.7
Ellipticity (λ = 222 and 282) of Trp Cage Variants Relative to WT

Construct	*Termini*	θ_{222} / θ_{222}**(WT)**	θ_{282} / θ_{282}**(WT)**
WT	**Free**	**≡1**	**≡1**
K8R		**1.46**	**1.63**
R16K		**0.895**	**0.675**
DM		**1.14**	**1.03**
WT	**Capped**	**2.11**	**1.42**
K8R		**3.63**	**1.12**
R16K		**1.47**	**1.46**
DM		**1.94**	**1.82**

Although Arg and Lys substitutions may stabilise the Trp cage fold to different extents, none of them annihilates it and it is safe to assume during further analysis that that fold is the most populated self-organised structure adopted by these constructs in solution at near-physiological conditions, as for our spray conditions of 10mM ammonium acetate (pH ≈ 6.8).

3.8.2 Collision-Induced Dissociation

The stability of N- and C-terminal capped Trp cage ions to vibrational excitation illustrates quite nicely the differences between proton affinities of the four point mutants. During CID in a quadrupole ion trap trajectories of m/z-selected ions are destabilised by means of an RF potential applied at the two end-cap electrodes, causing the beam to trail longer paths thus increasing the frequency of ion-neutral collisions as well as energy transfer during collisions. Collision energy depends on the mass and charge of the ions, the mass of the neutrals and is controlled by the magnitude of the RF amplitude; at a sufficiently high RF amplitude 'parent' analyte ion will dissociate into 'daughter' fragment ions. Monitoring parent ion intensity – normalised with respect to the total ion count (parent and all daughter ions) – as a function of RF amplitude yields parent ion depletion curves which are indicative of the structural stability of parent ions during vibrational (collisional) excitation. Dissociation studies of peptides with different gas phase proton affinities have led to the development of the mobile proton model for the peptide CID mechanism[35]. According to the mobile proton view, protons of cationic

peptides are likely to be sequestered by a small number of basic sites and solvated by a number of hydrogen bond acceptors given the molecule is sufficiently flexible to allow so. High-energy collisions may cause the proton to dislocate onto a number of proximal sites with lower gas-phase basicity; such an event triggers further rearrangement around the charge eventually leading to dissociation (see section 1.5.8 and references therein). Within such a framework it becomes clear that the proton affinity of a peptide's basic sites as well as spatial geometries the molecule occupies when vibrationally excited will influence both sensitivity to collision energy and the resulting dissociation patterns. Vibrationally-induced parent ion depletion of the four capped Trp cage constructs' $[M+2H]^{2+}$ and $[M+3H]^{3+}$ charge states (Figure 3.13) reproduce the qualitative trends predicted from the mobile proton model exaclty.

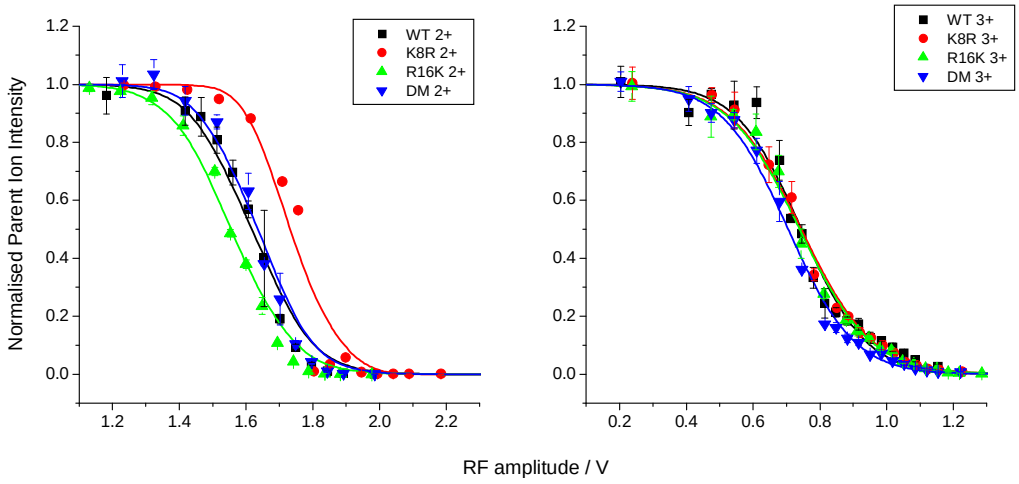


Figure 3.13
Normalised parent ion depletion curves for $[M+2H]^{2+}$ (left) and $[M+3H]^{3+}$ (right) ions for the four Trp cage constructs with protected termini. $[M+2H]^{2+}$ ions follow a trend that correlates well with the gas-phase basicity of the constituent amino acids; when a third proton is added these differences are abolished.

Stability of 2+ ions with collision voltage is well-correlated to the gas-phase basicity of their respective constituent amino acids (RR>KR>KK) leading to an order of stability to dissociation of K8R>WT=DM>R16K. Arguably, this finding supports the idea that the proton in the $[M+2H]^{2+}$ gas-phase ion is sequestered by B8 and B16 – at least after

collisional excitation which is an ergodic vibrational activation process – and that arginine, sporting a highly basic guanidino group, is more effective in doing so than lysine. Therefore alternative protonation site assignments may become prominent only if Trp cage ions remain conformationally rigid during desolvation, not allowing the protons to mobilise and eventually 'relax' to their most "intuitive" positions. This hypothesis also hints at deeper implications when studying protein conformation by MS: the temperature of protein ions from their native state in solution until they reach the detector and will have an effect on their conformation.

The correlation between predicted gas-phase proton affinity and stability to high-energy collisions observed for $[M+2H]^{2+}$ ions is not detected for $[M+3H]^{3+}$ ions. Once more, the mobile proton model can account for this phenomenon: since terminally protected Trp cage contains two basic sites the third proton will most likely occupy a range of quasi-degenerate[*] sites. Thus addition of a single proton to the $[M+2H]^{2+}$ state overrides the effect of the side-chain proton affinities. The presence of a mobile proton, in addition to a looser structure, also explains why $[M+3H]^{3+}$ ions dissociate at significantly lower (about half) collision voltages than $[M+2H]^{2+}$ species.

The trends observed could be taken to imply the stability of the low energy gas-phase conformation for each charge state – however this claim is fallacious. Under the conditions studied, peptides are excited to such an extent that fragmentation takes place. As each round of trapping is in the millisecond / second regime, even local perturbations of non-covalent structure mediated by collisions will have time to affect neighbouring loci. Furthermore, as excitation is sufficient to disturb covalent structure, rearrangements of non-covalent interactions can be readily assumed to take place at an energy range lower than that needed for dissociation.

Although gas-phase conformation cannot be directly inferred from CID data, measurements described here verify the order of proton affinity of the Arg- & Lys-

[*] "Quasi-degenerate" is defined here in the following terms: fluctuation of the peptide's (conformational) potential energy is similar or larger than the potential energy difference between the different protonation states considered. It is also understood that the hypothesis outlined in the main text can be verified only after extensive thermodynamic sampling followed by QM calculations for proton motion, yet the data themselves are quite convincing.

substituted Trp cage and provide experimental support for the decisions that were necessary to be made when considering proton localisation assignments for MM modelling (sections 3.5 and 3.8.5).

3.8.3 Hydrogen-Deuterium Exchange

Gas-phase HDX offers an attractive approach for probing properties of solvent-free ions forming higher-order structures stabilised by hydrogen bonding. HDX is achieved by including a deuteron carrier in the neutral dampening gas of an ion trap, as described in Chapter 2. Several mechanisms could account for the gas-phase HDX reaction, but for peptide based systems exchanging with D_2O or CD_3OD there is a consensus that the relay mechanism[36] is responsible. For HDX to occur via the relay pathway, the deuteron carrier must collide in such on orientation with the analyte ions so that a proton acceptor forms a hydrogen bond with a deuteron, and a proton on the analyte hydrogen-bonds to an acceptor on the D-carrier. HDX thus occurs almost simultaneously, upon dissociation of the complex. When CD_3OD is used as a D-carrier, as was done for the work described herein, the hydroxyl oxygen acts both as the D donor and H acceptor, setting a spatial constraint on the positions of the analytes D acceptor and H donor sites, whose interatomic distance must not exceed ~5Å [36]. Furthermore, if either of these sites participates in a hydrogen bond, HDX becomes less likely, lest that hydrogen bond be disturbed by the collision of the peptide with CD_3OD. Thus *in vacuo* D gain and H loss will be successful only if sites located on the peptide surface conform to an appropriate geometry, unlike the situation in solution, where the two processes are spatially (and temporally) uncorrelated, as nearby solvent molecules can often stabilise the D donor. Consequently, gas-phase HDX rates are generally lower than those observed in solution. Attention must be paid to the fact that a low HDX rate can be interpreted as either evidence for the existence of stable hydrogen bonding networks *or* the existence highly extended structure, not permitting the intermolecular interactions necessary for HDX to take place; additional data must be employed to differentiate between these two extremes.

HDX data were collected for the four capped Trp cage constructs over a time of 100s and were analysed using the first-order kinetic model for proton ($_1$H) loss described by Anderegg *et al.*[37]. Unfortunately trapping of ions for longer intervals resulted in signal loss, hindering measurements for 3+ ions; only $[M+2H]^{2+}$ species are considered herein.

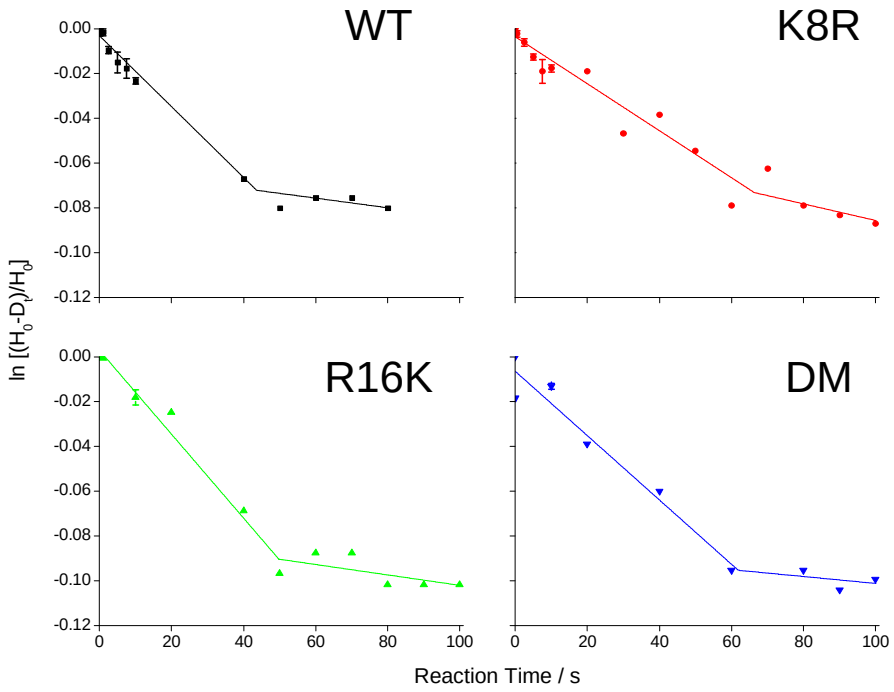


Figure 3.14
Gas-phase hydrogen-deuterium exchange kinetics for Trp cage constructs with protected termini up to 100s. Two exchange rates are obtained which indicate the presence of two classes of protons. The data analysis method outlined by Anderegg *et al.*[37] was followed.

Two features about the *in vacuo* HDX of Trp cage are immediately noticeable (Figure 3.14). First, all peptides exhibit a biphasic HDX behaviour. This feature has been observed for all peptide systems studied by the setup described in section 2.3.4.3 and is attributed to the different exchange rate of different protons[37]. All protons prone to exchange will do so early in the reaction; further exchange will depend on conformational fluctuations occurring along the peptide chain, exposing hydrogen atoms at geometries appropriate for HDX.

Table 3.8
Gas-phase hydrogen/deuterium exchange of 2+ charge states of capped Trp cage constructs {$[H]_0$: no. of non-aliphatic (exchangeable) protons; $[D]_{max}$: maximum (asymptotic) number of protons exchanged}.

Construct	$[H]_0$	$[D]_{Max}/[H]_0$	Rapid exchange rate / s^{-1}	Slow exchange rate / s^{-1}
WT	33	0.07 ± 0.01	$-1.58 \cdot 10^{-3} \pm 5.52 \cdot 10^{-5}$	$-2.15 \cdot 10^{-4} \pm 1.50 \cdot 10^{-4}$
K8R	35	0.08 ± 0.01	$-1.05 \cdot 10^{-3} \pm 9.47 \cdot 10^{-5}$	$-3.70 \cdot 10^{-4} \pm 2.69 \cdot 10^{-4}$
R16K	31	0.09 ± 0.01	$-1.87 \cdot 10^{-3} \pm 1.35 \cdot 10^{-5}$	$-4.26 \cdot 10^{-4} \pm 1.42 \cdot 10^{-4}$
DM	33	0.09 ± 0.02	$-1.33 \cdot 10^{-3} \pm 2.76 \cdot 10^{-5}$	$-1.55 \cdot 10^{-4} \pm 1.32 \cdot 10^{-4}$

Secondly, all Trp cage constructs are characterised by fairly slow exchange rates and equivalently low final number of exchanges, $[D]_{max}$ with only approximately three protons exchanging over 100s of trapping. It suffices to say, for a rough qualitative comparison, that 26-residue bee venom melittin $[M+3H]^{3+}$ ions studied under very similar experimental conditions as Trp cage ions exchange about 30% of their 52 exchangeable protons during the first 10s of the reaction[38]. Clearly, $[M+2H]^{2+}$ states of all four capped Trp cage constructs either possess a fairly robust structure or are completely extended. Given the compact nature inferred from IMMS (sections 3.4 & 3.8.4) and molecular modelling (sections 3.5, 3.6, 3.8.5 & 3.8.6) the latter option must be abandoned. All four constructs undergo a nearly identical number of exchanges at the long time limit, indicating that their gas-phase conformation is similar. Exchange kinetics differ subtly. Initial (rapid) exchange rates correlate roughly with proton affinities, supporting the hypothesis that charged protons are somewhat less localised in Lys-containing peptides. However, this trend does not persist to longer reaction times. Such a finding agrees with the view that ion population becomes redistributed over configuration space (including location of charged protons) when ions become thermalised (e.g. during desolvation or injection into the ion trap) but are stable in the trap in the absence of an activating external potential. In addition, all constructs exchange to essentially the same extent despite the fact that they contain different number of exchangeable protons, suggesting that B8 and B16 form multiple non-covalent interactions – as expected by the charge solvation model – thus being protected

from exchange. However, exchange at these sites may also be partially impeded by the high proton affinity of lysine and arginine residues.

Overall, HDX of Trp cage in a quadrupole ion trap supports the theory envisaging existence of stable high-order structural elements in the gas phase. These characteristics were investigated further by means of IMMS experiments.

3.8.4 Ion Mobility of Trp Cage Lysine and Arginine Mutants

So far CID and gas-phase HDX experiments revealed subtle differences in the behaviour of lysine & arginine Trp cage mutants. IMMS experiments can enable us to form a more coherent argument concerning the conformation of gas-phase Trp cage. Collision cross sections determined from drift velocities in helium for the eight constructs' 2+ and 3+ charge states, as well as their sodium adducts, can be seen in Table 3.9.

With the exception of K8R, differences between other constructs are small, lying below the fluctuation attributable to experimental error. All ions except capped 3+ charge states (and associated sodium adducts) are significantly more compact than the cross-section calculated for the NMR structure, suggesting that the Trp cage fold does not survive intact into the gas phase using this experimental setting, but undergoes a structural collapse. As will be discussed in section 3.8.6, MD simulations of capped $[M+3H]^{3+}$ WT ions indicate several rearrangements taking place when the Trp cage fold is allowed to 'relax' *in vacuo*.

General trends described in section 3.4 generally hold, with the exception of the effect of sodium binding. Sodium adducts yield slightly smaller cross sections (compared to the un-adducted protonated analogue) for WT (unprotected termini) $[M+2H]^{2+}$ and all constructs with protected termini $[M+3H]^{3+}$; the remaining eleven ions give rise to larger cross sections upon sodium binding.

Table 3.9
Collision cross sections (in $Å^2$) of Trp cage constructs at ambient temperature

Construct	*Termini*	$[M+2H]^{2+}$	$[M+H+Na]^{2+}$	$[M+3H]^{3+}$	$[M+2H+Na]^{3+}$
WT	Free	366.0 ± 20.2	361.9 ± 29.6	391.7 ± 11.3	408.7 ± 8.4
K8R		346.4 ± 17.3	385.9 ± 19.3	376.2 ± 18.8	396.0 ± 19.8
R16K		366.5 ± 18.3	374.0 ± 18.8	382.4 ± 19.1	407.5 ± 20.4
DM		385.7 ± 45.7	383.9 ± 19.2	378.1 ± 25.9	400.2 ± 20.0
WT	Capped	387.4 ± 12.6	396.5 ± 40.6	429.6 ± 38.2	426.6 ± 30.6
K8R		394.9 ± 11.5	414.7 ± 44.0	426.8 ± 33.4	421.0 ± 34.0
R16K		382.7 ± 5.4	388.8 ± 11.3	414.3 ± 16.4	400.5 ± 17.6
DM		381.0 ± 10.1	399.3 ± 15.7	424.1 ± 11.3	405.9 ± 16.9

The differences are small and may be attributed to experimental error. However, if they are deemed to be significant, then such a phenomenon would suggest that a Na ion disrupts a pre-existing stable conformation in order to allow better charge solvation in the said cases; yet in peptides that are relatively extended (3+ capped peptides) Na binding yields 'tighter' structures, presumably by recruiting backbone and/or side-chain carbonyl groups for its coordination.

Differences between peptides with protected and unprotected termini extend beyond their charge-dependent sodium binding-dependent properties. Trp cage ions with free termini are consistently more compact than their capped-termini equivalents. The only exceptions being DM $[M+2H]^{2+}$ & R16K $[M+2H+Na]^{3+}$ and in these two cases the difference between their respective cross sections is sufficiently small to have arisen accidentally. Undoubtedly this difference can be partially assigned to the mere presence of protecting groups (acetyl and amine). However capped and uncapped peptides also

differ in the relative strength of their intermolecular interactions (i.e. the strength of hydrogen bonding between a primary amine and a carboxyl group vs. that between two amides) as well as their gas-phase basicity (the unprotected N-terminus is more likely to become protonated than the acetylated equivalent). These experiments yield the first indication in this study that not simply charge but also proton location can play a crucial role in determining solvent-free peptide conformation.

K8R, the construct with the bulkiest side-chains, displays somewhat different behaviour to the other constructs. Capped $[M+2H]^{2+}$ K8R ions are more extended compared to other constructs; yet the opposite is true for free-termini K8R. As arginine residues have the highest proton affinity[33], it could be possible that when the termini are unprotected, a significant population of ions is zwitterionic, enabling the formation of a salt bridge and yielding more compact configurations.

As described already for WT, structural assignments for the experimental cross sections were attempted by comparing them with configurations generated *in silico*.

3.8.5 Gas-phase Configurations Generated by Simulated Annealing

Experimental studies so far have shown that of the eight constructs currently discussed, only differences detected by CID are clearly correlated to gas-phase basicity of constituent amino acid side chains. Gas-phase HDX and IMMS indicate a more complex behaviour with regards to conformation that is dependent on factors other than proton affinity. We proceed with a presentation of the theoretical studies on the constructs which, as for WT, consisted of simulated annealing and MD simulations.

Free-termini charge assignments studied by simulated annealing for all peptides are equivalent to these listed under Table 3.3 with the exception of 3b and 3c, which were simulated only for the original sequence, WT. An N-terminal acetyl cap reduces the basicity of the N-terminus significantly; therefore, for 3+ charge states of capped constructs alternative sites must be considered. Since, as discussed in section 3.5, the possibilities are quite varied when based on sequence alone, owing to the presence of several identical or near-identical amino acids in the sequence, only two possibilities were considered. The first is $Gln5^{+}$, as Adams *et al.* have argued for this scenario for the

free-termini peptide[27]. The second is $Pro12^{+}$; not only do proline residues possess a relatively high gas-phase proton affinity[33] but also Pro12 appears to be the most important site whose solvent-free configuration coincides with that of the solution conformation, singling it out from other proline residues. Furthermore, as Pro12 is located near the middle of the peptide chain it may be a stable site of self-solvation. Other assignments are possible but we will limit our analysis to these two. Charge assignments and the nomenclature of capped WT studied are listed in Table 3.10.

Table 3.10
Charge assignment of modelled peptides (capped termini)

Label	***Net Charge***	***Proton location***
WTc2	2+	NLYIQWL**K**DGGPSSG**R**PPPS
WTc3a	3+	NLYI**Q**WL**K**DGGPSSG**R**PPPS
WTc3b	3+	NLYIQWL**K**DGG**P**SSG**R**PPPS

Resulting geometries were analysed in the same way as the WT free termini peptide described in section 3.5. Same charge assignments of different constructs also resulted to very similar behaviour (data not shown). This finding can lead to suggesting that although the location of the charge plays a significant role in determining the gas-phase conformation, in systems like Trp cage differences between Arg and Lys become insignificant. A useful conclusion of this observation can be summarised thus: adoption of a coarse-grained modelling approach, speeding up calculations, is likely to adequately represent the more general conformational properties of such systems. However we must stress once again that the most likely difference between arginine- and lysine-containing peptides is found in the tendency for arginine to abstract the proton in low charge states. Therefore, lysine-containing variants are likely to occupy a *larger population of different charged sites*, whereas the set is more constrained in arginine-containing systems, analogous to their relative proton affinity and geometric rigidity.

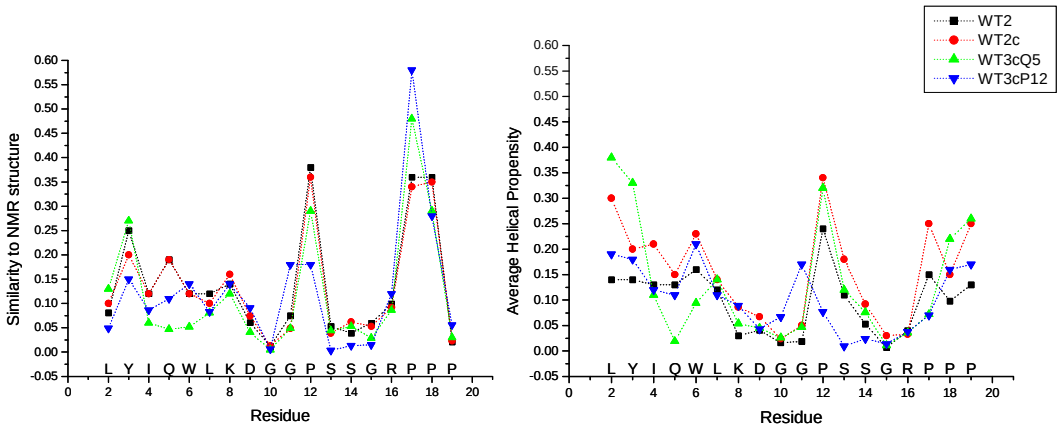


Figure 3.15
Similarity to NMR structure (left) and helical propensity (right), calculated using equation {3.1}, which normalises values to 1 (the latter value being attributed to the reference geometry), for three different charge assignments of WT containing protected termini. Differences between charge assignment are more significant than the differences displayed by K8R, R16K and DM (latter three not displayed).

As noticed in section 3.5 some essential structural traits are shared among all simulated molecules:

- Preservation of the conformation of proline residues, most notably Pro12, leading to formation of U-like topologies (further altered by additional turns)
- Helical turns occur mostly in the N-terminal segment (but are also present over residues 17-19)
- Gly10 and Gly15, which in essence control the packing of different secondary structural elements, almost never adopt their solution configuration.

Since differences between charge assignments are of greater significance than the exact structure of the charged amino acid, data for the three additional constructs of the original Trp cage sequence (WT) are displayed in Figure 3.15. Comparison with Figure 3.5 reveals the effect of terminal capping groups on the structure. The N-terminal helix of 2+ capped ions is stabilised compared to 2+ ions with free termini. Presence of a proton on Gln5 side-chain has a similar effect in the capped peptides as with the free termini constructs; structure around Gln5 is perturbed with concurrent stabilisation of a helical configuration of the upstream sequence. Locating the third proton on Pro12

destabilises native-like characteristics at residues 12-15; yet the proton on Pro12 forms a relatively persistent interaction with Gly11 which assumes a more helical character. Thus, as inferred from Figure 3.15, the charged proline has a different conformation than a neutral one, achieving a translocation of the central turn by one amino acid.

Interestingly, relative stabilisation of the N-terminal helix in the capped constructs agrees with CD data.

Calculated cross-sections for the 2+ charge state of the capped peptide are similar to that of the free-termini peptide, with one subtle difference: distribution of cross sections is broader for the capped constructs, with both more extended and more compact states being sampled than in their free termini counterparts. 3+ charge states of capped constructs on the other hand are uniformly more extended than free termini 3+ ions. From this observation we can infer that the hydrogen bond between the termini is decisive in stabilising more compact structures at higher charges, when repulsive forces are more pronounced.

These trends can be discerned in Figure 3.17, depicting exponentially-weighed average cross-sections of the four constructs. Capped 3+ charge states of K8R and R16K are unusually large compared to the remainder of the peptides and experimental values. Although such an event may be considered as evidence against the simulated charge assignments, attention must be drawn to the fact that these values result from only a handful of simulated structures and statistical convergence issues may come into play.

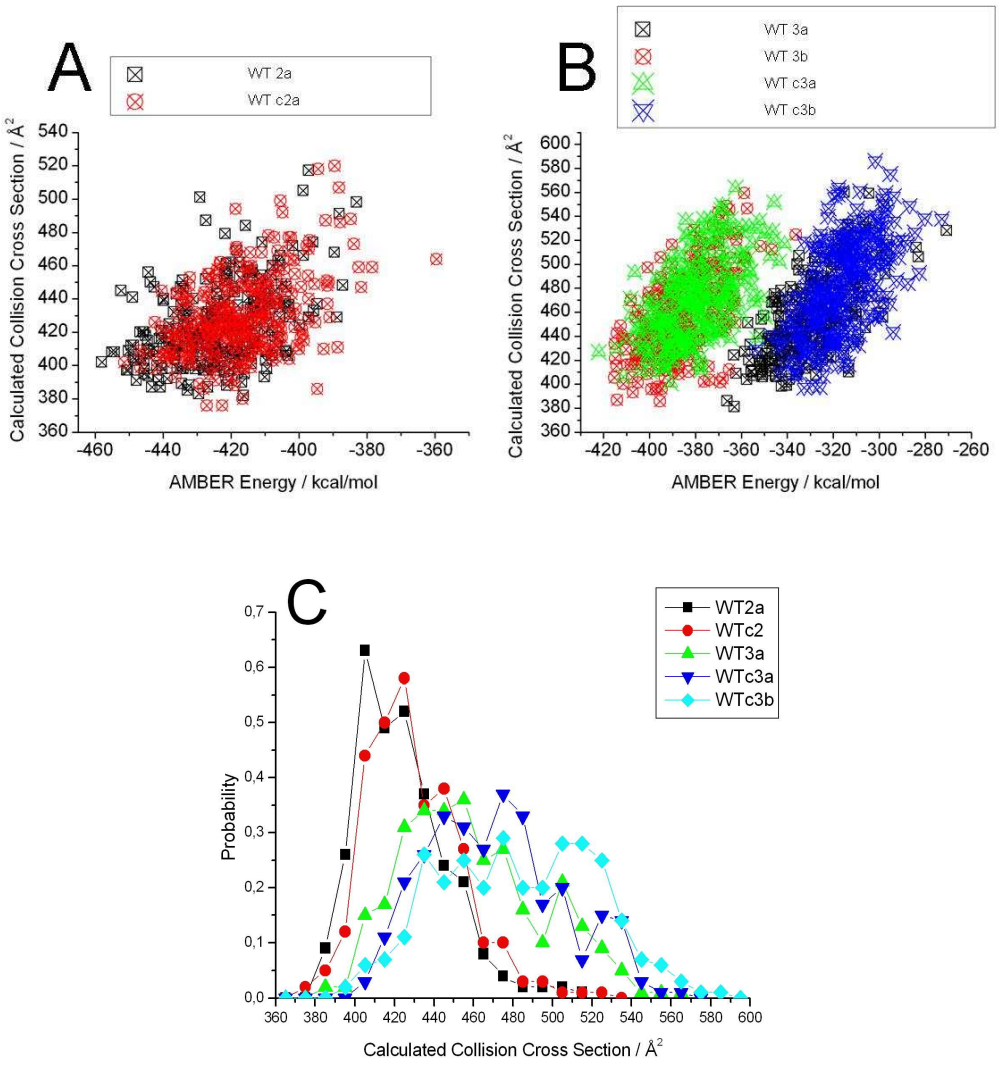


Figure 3.16
Comparison of cross section vs. energy scatter plots between constructs with protected and unprotected termini. A) $[M+2H]^{2+}$; B) $[M+3H]^{3+}$; C) Histograms of collision cross sections for five constructs, indicating illustrating the effect of charge and capping groups on overall Trp cage structure (Graphs A and B are not in scale).

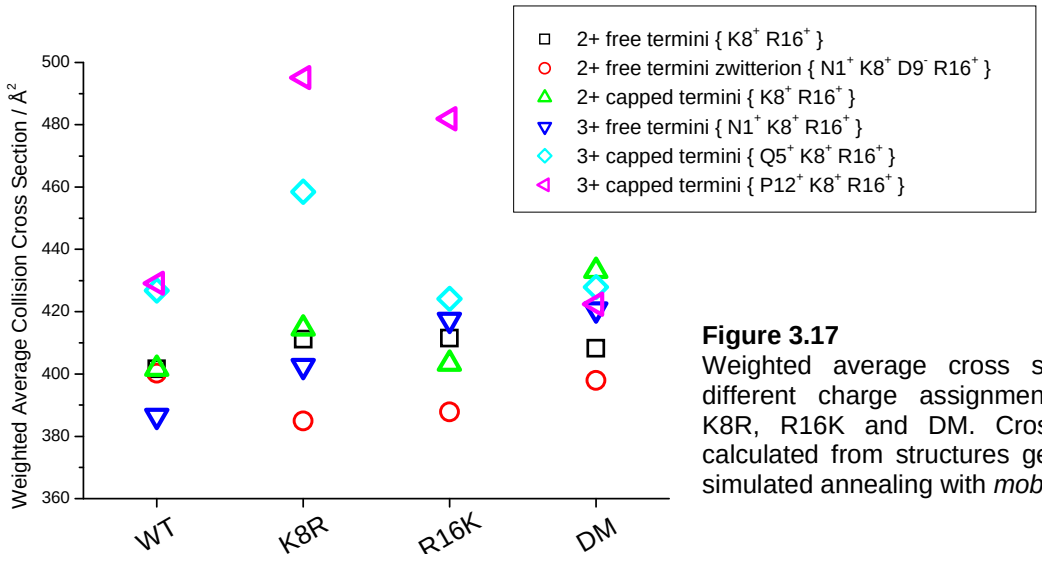


Figure 3.17
Weighted average cross sections for different charge assignments of WT, K8R, R16K and DM. Cross sections calculated from structures generated by simulated annealing with *mobcal*.

A direct comparison between calculated cross-sections displayed in Figure 3.17 and experimental results shown in Table 3.9 can be seen in supplementary figure App.D.1. Collision cross sections estimated from simulated annealing structures are generally larger than the ones obtained experimentally. The discrepancy is more pronounced for the free-termini constructs. This effect, as already mentioned, can be either due to instrument parameters whose effect was not taken into account or imperfect empirical parameters used in the simulation. This discrepancy becomes most evident for 2+ charge states of peptides with unprotected termini but is also evident, to a lesser extent, in the case of unprotected $[M+3H]^{3+}$ ions (apart of WT 3+) and protected $[M+2H]^{2+}$. If however the data are assumed to be comparable, cross sections calculated for $[M+2H]^{2+}$ zwitterions fall closer to the experimental data for peptides with unprotected termini. Simulated cross sections for $[M+3H]^{3+}$ ions fall closer to experimental data, especially in the case of peptides bearing protecting terminal groups; for the latter calculated cross sections for all constructs apart of K8R, the third proton located on the Q5 side-chain, agree fairly well with experimental values. That convergence of K8R may be more problematic than for other constructs can be understood; not only does K8R contain more atoms than all other constructs, but the structural rigidity of arginine side-chains can be envisioned to impede their coordination, especially as two arginine residues

interacting with the peptide are bound to come in close proximity. Yet the rescaled cross-sections reproduce many qualitative results seen experimentally, namely: the larger cross sections yielded for capped K8R than other capped constructs; capped constructs being generally more extended than peptides with free termini; addition of a charged proton yielding larger cross sections. Thus, as far as these more general traits are concerned, this methodology is capable in portraying them; however, more detailed structural studies require more exhaustive and more sophisticated conformational sampling algorithms. Under the assumption that native-like characteristics can endure the electrospray process and mass spectrometric analysis, study of trajectories beginning from a configuration representative of the solution structure may be more appropriate. Such MD simulations constitute the next step of the analysis.

3.8.6 Molecular Dynamics of Lysine-Arginine Mutants at 300K

Gas-phase MD trajectories were calculated for the three 'mutants' studied experimentally (K8R, R16K and DM) as well a D9N construct which was only treated computationally. The D9N construct, results for which have been reported elsewhere by different methods[14, 29, 39], does not form a K8-D9-R16 salt bridge and if some synthetic peptide can be obtained in the future it will certainly constitute an interesting system for comparison. Whereas WT was treated for both free and protected termini, only peptides with unprotected termini were simulated in all other cases. As the fundamental characteristics of the trajectories are shared between all constructs and differ only with charge assignment, our analysis does not suffer significantly from the absence of such data. Charge assignments simulated for the mutant sequences were the equivalents of WT2, 2z and 3a, using the nomenclature explained in Table 3.3.

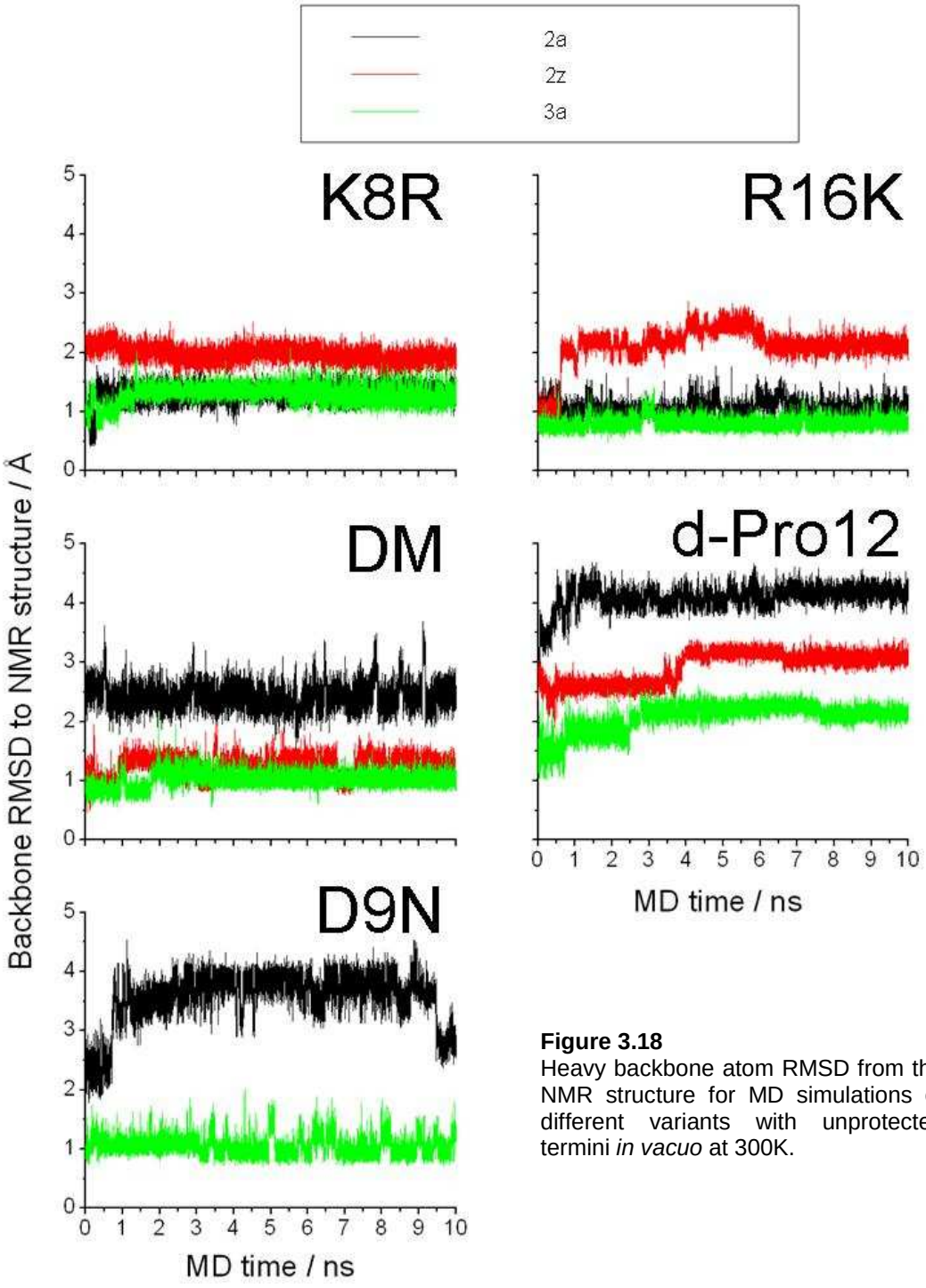


Figure 3.18
Heavy backbone atom RMSD from the NMR structure for MD simulations of different variants with unprotected termini *in vacuo* at 300K.

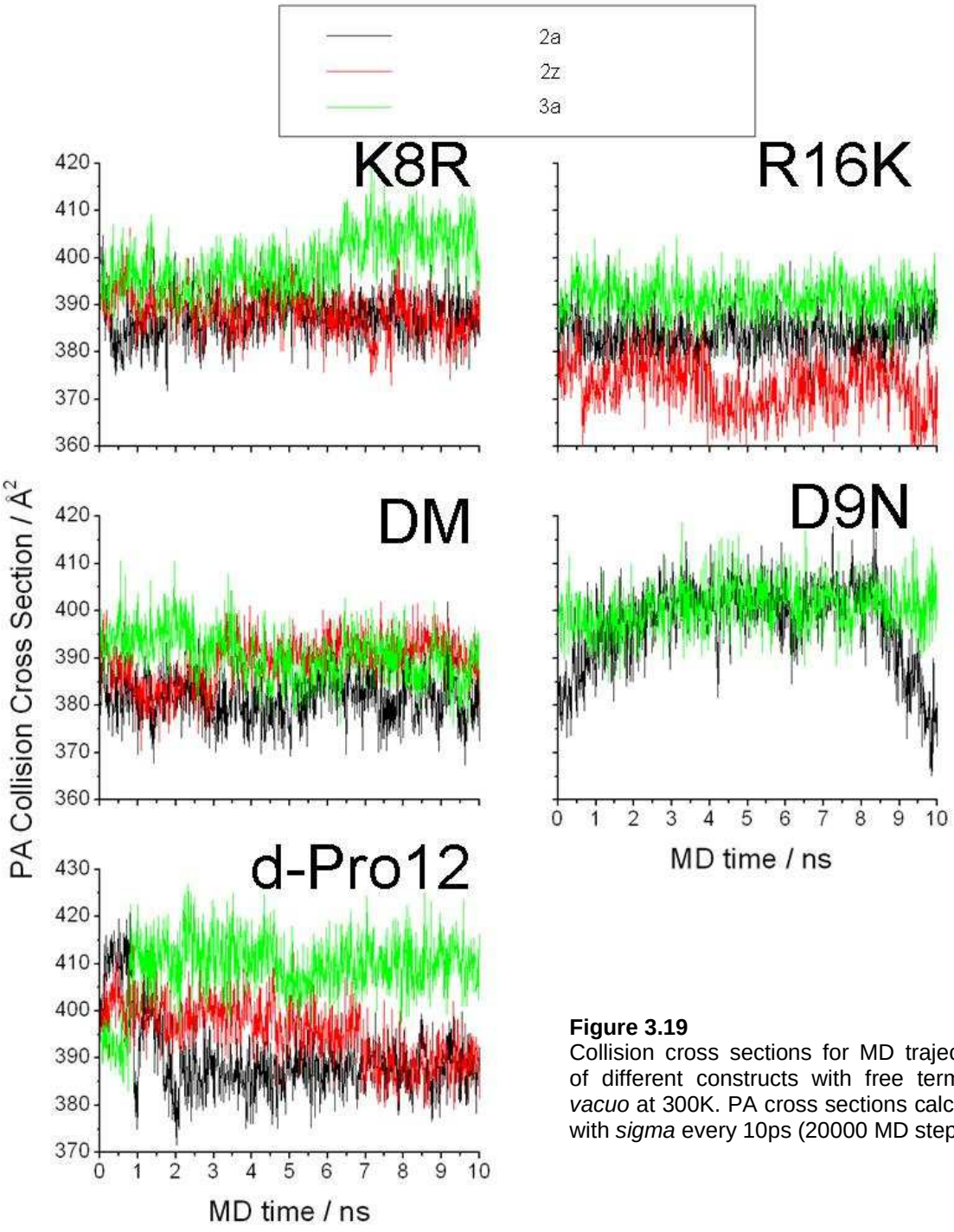


Figure 3.19
Collision cross sections for MD trajectories of different constructs with free termini *in vacuo* at 300K. PA cross sections calculated with *sigma* every 10ps (20000 MD steps).

MD trajectories at 300K for different constructs are summarised in Figures 3.18 and 3.19, displaying heavy-atom backbone RMSDs to the NMR structure and corresponding *sigma* cross-sections respectively. General features observed by simulations of WT (section 3.6) are shared amongst *the majority* of simulations, namely:

- Large-scale conformational rearrangements occurring at low temperatures relax the peptide into more compact structures – i.e. dehydration results into structural collapse driven by self-solvation
- 2+ zwitterionic constructs are stabilised by means of a salt bridge
- Non-zwitterionic 2+ show greater fluctuations, due to coulombic repulsion as well as the absence of strong electrostatics between the termini
- The solution fold in simulated 3+ charge states becomes stabilised by the ion-dipole interaction between the terminal groups and displays the smallest fluctuations at 300K.

Several exceptions to trends established by analysis of WT trajectories, as well as intriguing differences between constructs are also apparent. K8R 3a preserves a backbone topology very similar to that of the NMR structure yet side-chain fluctuations give rise to more extended configurations. In contrast with WT2, both K8R 2 and R16K 2 preserve the solution fold to a greater extent than the 2z states. All simulated charge states of K8R, R16K and DM maintain recognisable features of the Trp cage backbone topology; yet those with the least similarity (K8R 2z, R16K 2z, DM2) give rise to more compact structures, supporting the hypothesis envisaging collapsed equilibrium gas-phase conformations. The effect is more pronounced in the case of R16K when compared to other constructs, indicating that it is indeed electrostatics that underlie gas-phase collapse. Furthermore, the charge on the lysine side-chain amine group is more localised compared to the equivalent guanidino group of arginine, explaining the difference between simulated K8R and R16K zwitterions. In qualitative agreement with experiment, 3+ peptides routinely yield larger calculated cross-sections than 2+ peptides. Importantly, the trends observed for K8R, R16K and DM are not obeyed by D9N. D9N 3a preserves a ‘native-like’ backbone topology at 300K; the resultant cross-sections

compared to other 3a constructs is somewhat larger. Additionally, although D9N 2 is *on average* more compact than D9N 3a, it is also much more extended than the other three 2+ constructs at 300K. Such observations confirm the importance of strong electrostatic interactions; hydrogen bonding mediated by N9 in the case of D9N are weaker than these formed by the carboxyl of D9 in all others, resulting in 'looser' structures.

MD simulations of mutant sequences at 300K indicate that conformational changes lead to the formation of 'non-native' but nevertheless compact gas-phase folds. Experimentally produced K8R ions might be more prone to forming zwitterions due to the high proton affinity of arginine. Optimal gas-phase conformations of zwitterions have been shown to be compact by simulated annealing (Figure 3.17). However if the presence of a zwitterion stabilises the solution fold, resulting ions will give rise to larger cross sections; finally, bulky Arg side-chains can further accentuate such an effect.

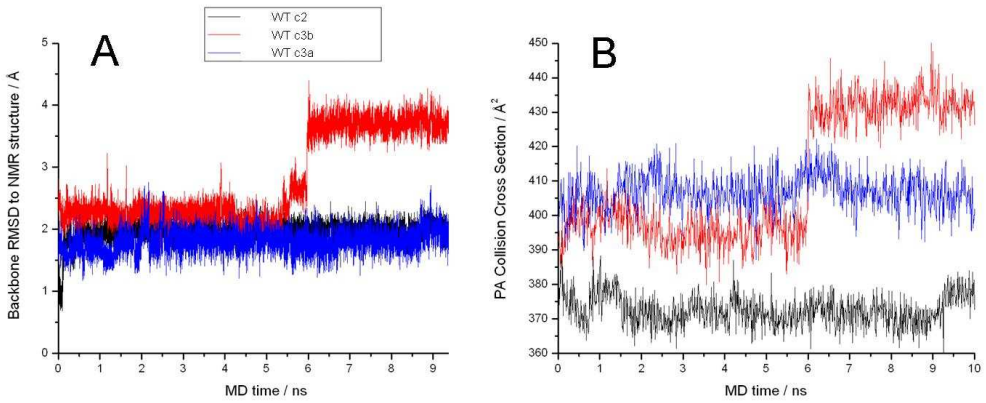


Figure 3.20
Backbone RMSDs to the NMR structure (left) and collision cross sections (right) obtained from MD simulations of Trp cage with protected termini.

To complement experimental data, MD of three WT constructs with protected termini were also run. Heavy-atom backbone RMSD and PA collision cross sections calculated with *sigma* for the three trajectories at 300K are displayed in Figure 3.20. Results for these constructs are in general comparable to these obtained for peptides with neutral termini. WTc2 undergoes a collapse at room temperature but still preserves a backbone topology that is recognisable as the Trp cage fold. WTc3a also preserves a native-like topology during the entire simulation at 300K; however, WTc3b undergoes a structural

change during the 5th nanosecond of the run. Interestingly the resulting structure is more extended than the previous configuration. Placing a proton of the backbone carbonyl of Pro12 has a destabilising effect on the solution fold. The simulations are in qualitative agreement with experiment, with both 3+ systems giving rise to higher cross-sections than the 2+ peptide. Notably, WTc3a,b are more extended than WT3 peptides (with unprotected termini) whereas WTc2 cross sections are comparable with these obtained for WT2a,b and the resulting gap between 2+ and 3+ capped ions is greater.

3.8.7 Effect of Lysine and Arginine Substitutions

Studies on K-R substituted Trp cage variants illustrate the influence of side-chain gas-phase basicity on CID fragmentation propensity, gas-phase HDX kinetics and peptide conformation. Although average collision cross-sections in helium were similar among peptides only differing in the structure of their basic residues, CID and HDX data exhibited significant differences. Consistent with the mobile proton model, dissociation energy and initial HDX rates of $[M+2H]^{2+}$ ions correlate with proton affinity of constituent amino acids, indicating that these phenomena proceed via proton transfer reactions; furthermore the protons exchanged initially in the gas phase are likely to be the charged protons. Interestingly, the proton affinity of the isolated side-chains also correlates with the stability of secondary structural elements in an aqueous buffer pH≈6.8 according to far-UV CD experiments (in the case of constructs with unprotected termini it also correlates with the stability of tertiary structure as shown by near-UV CD). Ionic mobilities, slow HDX rates and intensity of near-UV CD spectra for peptides with protected termini are not correlated with gas-phase basicity. In solution, the trend leads to the hypothesis that the stabilisation of secondary structures by addition of capping groups compensates for the stabilisation afforded by adjusting the proton affinities of the two basic residues and, consequently, the strength of the X8-D9-X16 salt bridge. The steric rigidity of guanidine may also disfavour long-lived interactions of arginine side-chains to D9. In the absence of solvent, we observe that slower exchange rates follow a more complex pattern. Two interpretations may be brought forward:

I. slow HDX kinetics depend on proton transfer from neutral groups to form short-lived zwitterions[40]
II. over the multi-second trapping timescales configurational changes may expose exchangeable sites.

As the basicities of functional groups in peptides are affected by their environment[33, 41] these interpretations do not oppose each other.

Addition of one proton over the number of basic groups abolishes the differences between constructs measured by parent-ion decomposition, indicating that that third proton is more 'mobile'. This effect is also reflected in the different cross-sections obtained for $[M+3H]^{3+}$ ions compared to $[M+2H]^{2+}$ ions, as already discussed. Simulated annealing and molecular dynamics at 300K simulations did not reveal any major differences between the constructs; however, differences between charge state and charge location are significant. Now a question which lies in the core of the correlation between CID and IMMS data, leaving instrumental differences aside is whether the activation energy needed to mobilise the proton in the 2+ charge state is sufficient to impart global conformational modifications to the geometry of peptide ions. Data on temperature dependence of desolvated peptide ion conformation, discussed in Chapter 4, may hint towards an answer.

3.8.8 Effect of Capping Groups

"Protecting" the peptide termini by incorporation of an acetyl group at the N-terminus and an amine at the C-terminus (so that both termini contain amides) directly influences the proton affinity of these groups[33] as well as modifying the strength of hydrogen bonding between the termini. These observations tie in fairly well with experimental and simulation results. The differences in the cross-sections of between peptides with free and capped termini are clear by IMMS. Capped constructs give rise to higher collision cross-sections in both 2+ and 3+ species, a result that can either be interpreted as the existence of 'looser' structures *or* greater persistence of solution-like conformational characteristics. In either case, experiments and simulations described herein implicate the N-terminus as a charged site in the case of free-termini peptides, and show that a

strong interaction between the peptide termini drives a structural collapse in the absence of solvent over time-scales and energies studied. Interestingly, MD simulations at 300K indicate that protonation of the N-terminus exerts a structure-stabilising effect. The importance of such interactions is discussed even further in Chapter 4.

The acetyl group at the N-terminus exerts a helix-stabilising effect, which is witnessed directly in solution by far-UV CD measurements. *In vacuo* the same situation results for no competition can exist in the gas phase between intramolecular hydrogen bonding and interactions with solvent molecules, and protonation at the N-terminus destabilises helical structures[32]. Simulated annealing also revealed a greater helical propensity for capped constructs near the N-terminus, supporting the validity of these statements. Thus, even though gas-phase basicity of X8 and X16 exerts only a minor effect on global structure at room temperature, the basicity of the termini seems to be detrimental.

3.9 Studies of a peptide construct containing a d-amino acid

Protein folding depends on the relative packing of the polypeptide. The chiral nature of 19 common amino acids (Gly being achiral) invites the incorporation of amino acid enantiomers as a method of inflicting a localised structural modification and monitoring its effect on the overall conformational characteristics. As studies of Trp cage presented here aim in part to outline the sensitivity of gas-phase techniques to protein conformation, studies of a peptide construct which is not expected to form stable 3° structure in solution can contribute significantly towards this purpose. Constructs containing d-amino acids have been previously studied[25, 27]; although differences in the observed ECD fragmentation pathways for d-substituted constructs were reported, replica exchange simulations *in vacuo* also revealed significant similarities in the conformations of these constructs. d-Tyr3, d-Gln5 and d-Leu7 are modifications chosen to perturb both secondary and tertiary structure as Tyr3 is situated in the N-terminal helical stretch and π-stacks to Trp6 in the NMR structure, Gln5 may interact with Lys8 and Leu7 also forms part of the miniprotein's hydrophobic core[23]. As it was deemed that a mutation effecting tertiary structure but leaving the most prominent secondary

structural motifs intact would be interesting to investigate as such a construct could indicate whether 3° structure stabilises 2° structure. The d-Pro12 construct was therefore designed. Pro12 constitutes part of the 'cage' to which Trp cage owes its name by forming CH-π interactions with Trpϐ[23]. Moreover, Pro12 does not belong to neither the N-terminal helical region nor the C-terminal PPII region but, as discussed in sections 3.5 and 3.8.5, it plays a pivotal role in determining the overall topology of the peptide, even in the gas phase. These characteristics render it an interesting variant for study. We may begin our investigation by observing the effect of the d-Pro12 substitution in solution, probed by CD spectroscopy.

3.9.1 Conformational Characteristics in Solution

Figure 3.21 displays far- and near- UV CD spectra of d-Pro12 along with WT for comparison. d-Pro12 gives rise to a far-UV spectrum commonly associated with 'random coil' peptides, exhibiting a single minimum around 203nm. Clearly, the 222nm band associated with α-helices is not present in the d-Pro12 spectrum. Additionally none of the characteristic bands are observed in the near-UV spectrum, which shows great resemblance to that of WT collected at 80°C (the thermal stability of the Trp cage fold is discussed in the next chapter).

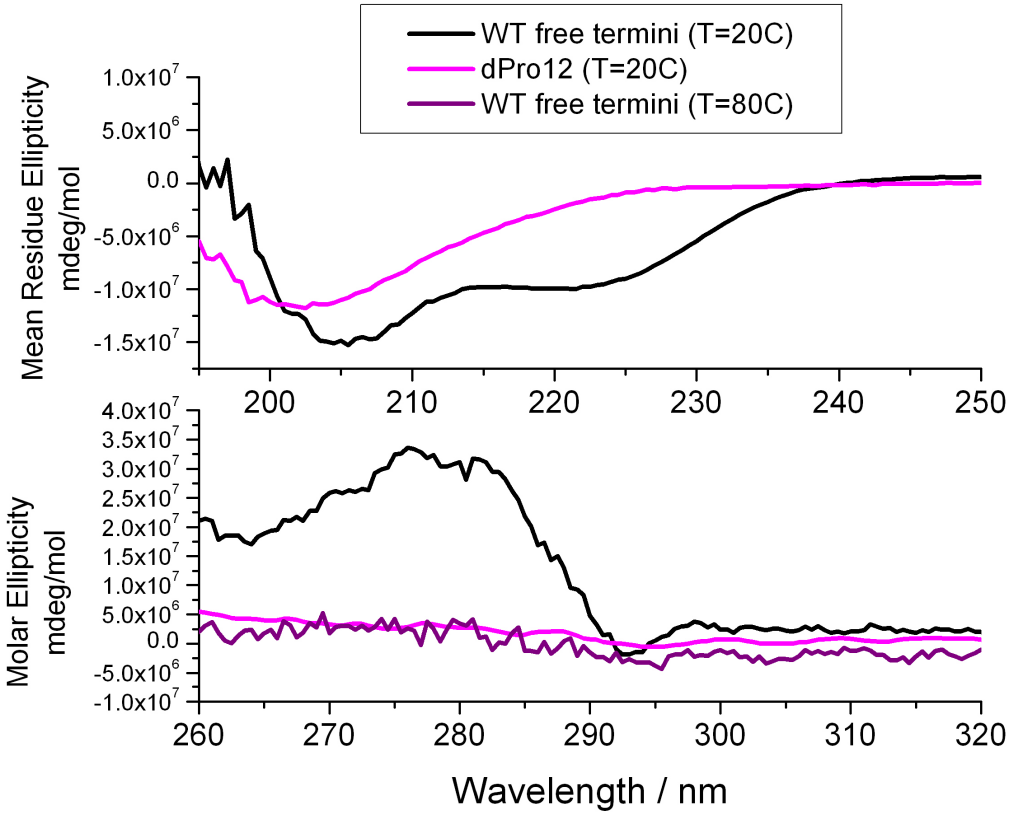


Figure 3.21
Far (top) and near (bottom) UV CD spectra of d-Pro12, compared with the enantiopure WT. Notably, the near-UV spectrum of d-Pro12 at 298K is very similar to that of WT at 353K.

One may, therefore, infer that the d-Pro12 construct does not give rise to any particularly persistent ordered non-covalent structure in an aqueous environment. The destabilisation of the N-terminal helix effected by d-Pro12 incorporation is of particular interest. As position 12 is not included in the helix, this observation suggests that the Trp cage fold is highly cooperative and the persistence of local structure is driven by the stability of the overall fold. These findings support the view envisaging hierarchy[42] in protein folding to be possible only in multi-domain proteins and not to hold in the case of small single-domain proteins. In the latter systems, formation of 2° and 3° elements proceeds concurrently. Further discussion on the mechanism of Trp cage fold formation is presented in the Chapter 4.

3.9.2 Mass Spectrum of d-Pro12

The appearance of the ESI mass spectrum of d-Pro12 (Figure 3.22) reflects the peptide's conformational characteristics in solution. As for all other constructs, 2+ and 3+ charge states are observed, along with some limited aggregation. However, the intensity of the $[M+3H]^{3+}$ charge state relative to that of 2+ ions is greatly increased compared to WT. This shift of the charge-state distribution towards the higher charge state indicates the existence of more extended conformations[43]. In addition to this shift, fragments resulting from hydrolysis of the peptide are much more intense in the case of d-Pro12. We may therefore conclude that presence of a stable compact fold protects WT from hydrolytic breakdown.

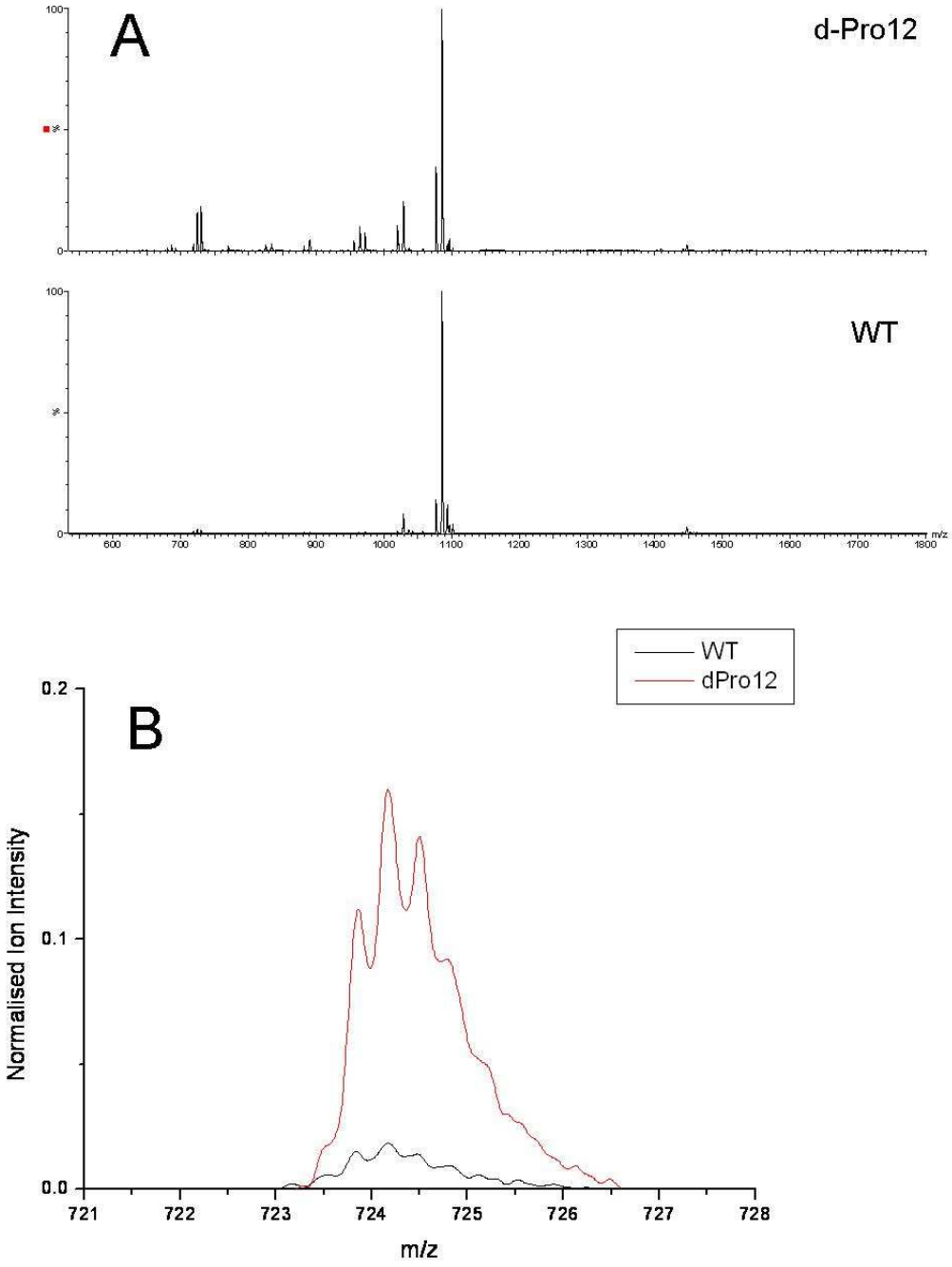


Figure 3.22
Mass spectra of d-Pro12 and WT collected using similar instrument settings. A) Entire spectrum, indicating the shift of ion population towards 3+ charge state as well as the greatest abundance of hydrolytic fragments in the case of d-Pro12. B) $[M+3H]^{3+}$ peak, normalised to the intensity of the $[M+2H]^{2+}$ peak. Compared to WT, d-Pro12 $[M+3H]^{3+}$ species is ca. 10 times more populated.

3.9.3 Ion Mobility of d-Pro12

Momentum-transfer cross-sections of d-Pro12 Trp cage ions are very similar with those obtained for the enantiopure peptide, as seen in Table 3.11. d-Pro12 ions yield only slightly higher cross-sections. This finding has several implications for the conformational stability of the Trp cage miniprotein in the absence of solvent as well as the sufficiency of IMMS data for determining gas-phase conformation. The two peptides clearly adopt very different configurations in solution under the conditions studied, as shown by CD (Figure 3.21). We may infer from these results that d-Pro12 collapses into compact, 'self-solvated' structures upon removal of solvent; it may even be possible that helices, which are less stable in solution, may form *in vacuo*. Finally, it is important to note that the internal energy of nano-ESI-produced peptide ions is not known exactly. The source temperature is kept constant at 80°C throughout all experiments; yet the capillary is not heated and the peptide solution should remain at ambient temperature. Furthermore, solvent evaporation from droplets produced by ESI lowers the temperature of the droplet[44]. Depending of the temperature of the ions and the thermodynamic stability of the Trp cage fold, even the original sequence may produce partially unfolded ions by nano-ESI. The thermodynamic stability of Trp cage in solution and in the gas phase is discussed in Chapter 4; now, simulations of the heterochiral peptide and the resulting interpretation of IMMS data will be described.

Table 3.12

Collision cross sections of d-Pro12 and WT at ambient temperature. Both experimental and theoretical cross sections are displayed. Theoretical values correspond to weighted averages based on simulated annealing structures. 2+ zwitterion theoretical cross sections are shown in italics.

	WT			**d-Pro12**		
Ion	*Experiment*	*Simulation*		*Experiment*	*Simulation*	
$[M+2H]^{2+}$	366.0 ± 20.2	401.7	*400.2*	378.1 ± 18.9	411.5	*393.6*
$[M+H+Na]^{2+}$	361.9 ± 29.6	---		362.6 ± 18.1	---	
$[M+3H]^{3+}$	391.7 ± 11.3	386.6		395.8 ± 19.8	417.1	
$[M+2H+Na]^{3+}$	408.7 ± 8.4	---		412.2 ± 20.6	---	

3.9.4 Simulated Annealing in vacuo

In contrast to the situation in solution, WT and d-Pro12 constructs only display minor differences as probed by IMS. The situation of the gas phase ions is analogous *in silico*, as probed by simulated annealing. As illustrated in Figure 3.23, incorporation of a d-Pro at position 12 only has a local effect on the simulated peptides' structure. For all three charge assignments of d-Pro12 examined, d-Pro12 constructs yield similar data over the entire peptide sequence, with the exception of residues 10-12.

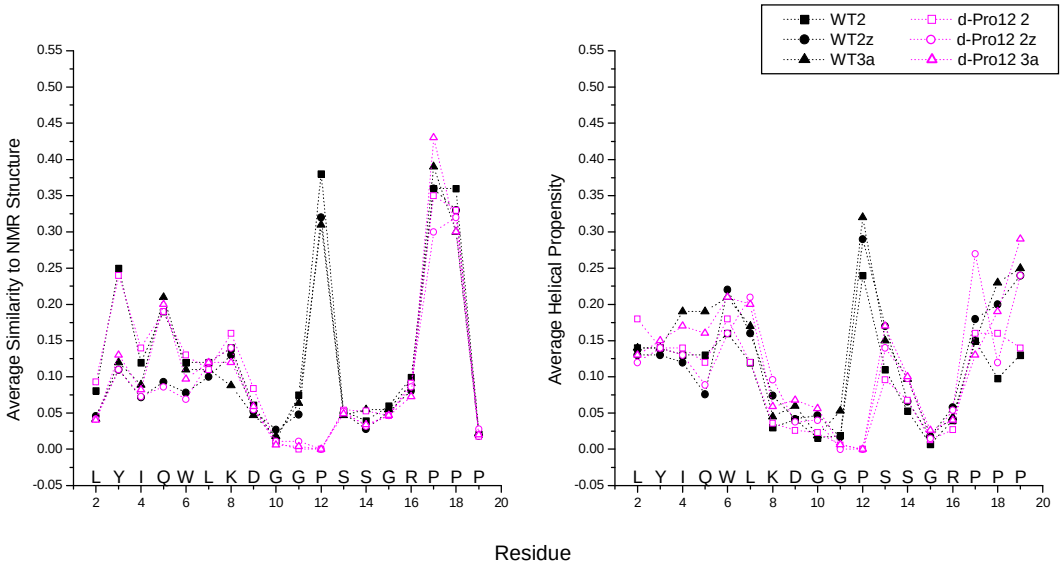


Figure 3.23
Similarity to NMR structure (left) and helical propensity (right) per residue for d-Pro12, calculated from geometries generated by simulated annealing. Analogous data states of WT are also shown for comparison. 'Propensities' calculated with equation {3.1} which normalises values to 1 (the latter value being attributed to the reference geometry)

The similarity extends to the weighted average theoretical cross-sections, as can be seen in Table 3.11. These observations can be summoned to reach several conclusions concerning the gas-phase properties of these peptides. *In vacuo* dispersion interactions between apolar residue side-chains contribute little to the stability of the overall structure; furthermore the absence of solvent renders self-solvation the main driving force determining peptide configuration. Hydrogen bonding between water molecules and the peptide may disorganise the secondary structural elements, which – when solvated – depend on the presence of hydrophobic tertiary contacts in order to become

long-lived. When no water is around, local structure will form the most stable thermodynamically and kinetically accessible configuration: incorporation of d-Pro12 does alter the topology of the peptide, even though self-solvation enables the generation of compact structures, giving rise to cross sections that are indistinguishable from other Trp cage sequences. As seen in Figure 3.24, d-Pro12 peptides are extended around Gly10-dPro12, yet two additional turns form, immediately before and after this region, facilitated by the three glycine residues, so that the peptide is stacked against this central portion of the sequence. As observed for all simulated annealing geometries of protonated Trp cage peptides, charged residues, namely K8, R16 and the N-terminus (the latter protonated only in the $[M+3H]^{3+}$ state), form multiple interactions, stabilising the sampled low-energy configurations and underlining the significance of charge solvation in determining peptide gas-phase structure.

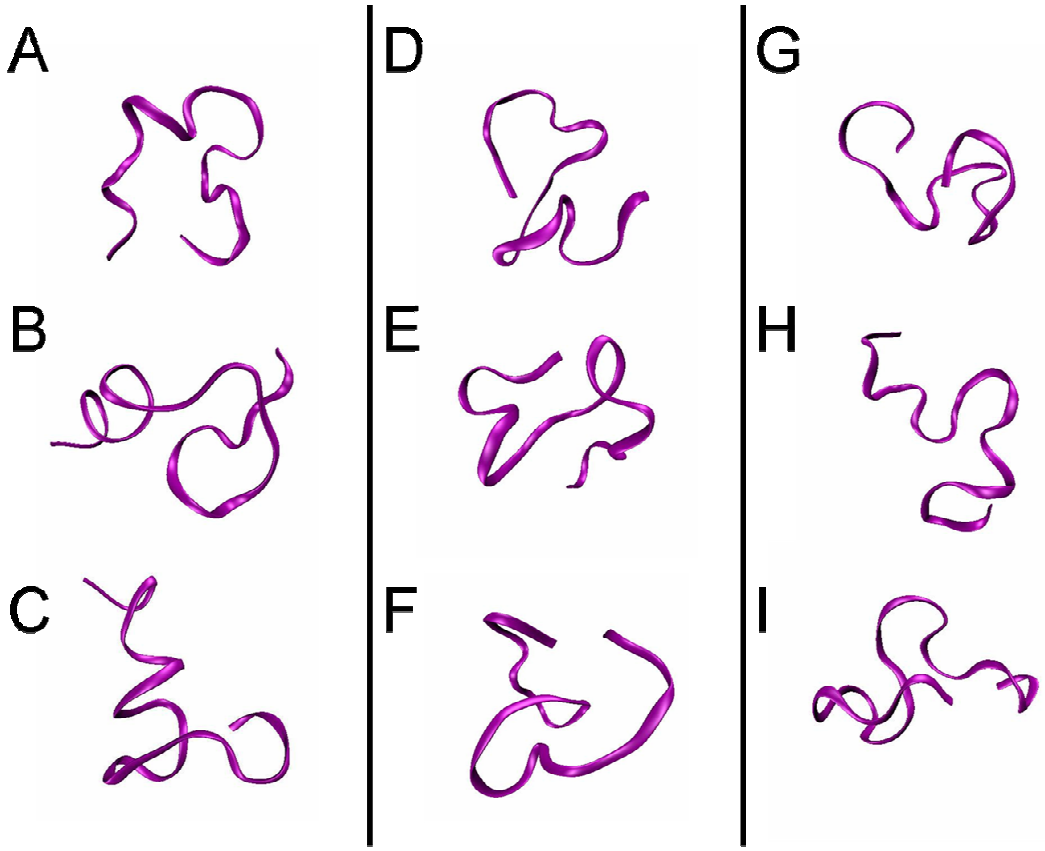


Figure 3.24
Low-energy structures of d-Pro12 generated by simulated annealing. A-C) d-Pro12 2a; D-F) d-Pro12 2z; G-I) d-Pro12 3a.

3.9.5 MD Studies

Order parameters for the d-Pro12 trajectories initiated from a fictive native-like fold (differing from 1L2Y only in the orientation of Pro12 side-chain so as to contain the desired chirality) are plotted in Figures 3.18 and 3.19. Compared to all other constructs, d-Pro12 has the most destabilising effect on the native fold, elements of which being sustained in the simulated 3+ charge state, but less so in the 2+ charge states. d-Pro12 2a undergoes significant rearrangements early on whereas in the case of d-Pro12 2z such events are delayed, which is indicative of the salt bridge's ability to "lock" the peptide into certain configurations, as indicated by backbone RMSD values. d-Pro12 3a gives rise to the highest cross sections of all three systems simulated; d-Pro12 2a collapses into more compact structures after an initial intermediate stage; d-Pro12 2z also becomes gradually more compact albeit at a slower rate than d-Pro12 2a. Comparing PA cross sections for d-Pro12 and WT (Figure 3.11; Table 3.11) we find that they are comparable; d-Pro12 is larger by *circa* 10Å^2 yet such a difference is well within the experimental error obtained for these values (sections 3.4 & 3.9.3). Thus the differences between these two classes of constructs cannot be based on ambient-temperature collision cross sections alone; yet a trace of their difference can be seen in the mass spectrum (section 3.9.2) and is also manifest in the more subtle aspects of the simulation data.

3.9.6 Effect of d-Pro12

d-Pro12 destabilises the Trp cage fold both in solution (as inferred from CD experiments and the appearance of resulting mass spectra) and *in vacuo* (as concluded from MD simulations). In aqueous solution, the presence of d-Pro impedes the formation of the N terminal helix as well as tertiary structure, illustrating the interplay between different 'levels' of structural organisation in the Trp cage miniprotein. However, IMMS experiments show that both the heterochiral and enantiopure construct adopt similarly compact geometries in the absence of solvent. This result is qualitatively replicated by structural optimisation *in vacuo* using simulated annealing in the case of $[M+2H]^{2+}$ ions but not for $[M+3H]^{3+}$ ions. Gas phase ions can be envisaged forming networks of intramolecular interactions that eventually perturb the packing characteristics of the

'native' structure. Evidence for such conformational transitions is given by the fact that experimental collision cross sections are much smaller than those expected for the NMR structure.

The absence of water abolishes the importance of hydrophobic interactions in stabilising non-covalent structures. Forces of self-solvation take the upper hand and, while retaining some native-like characteristics such as secondary structural elements, seek to maximise strong intramolecular interactions, eventually rendering the differences between d-Pro12 and WT undistinguishable. However, the different topologies of WT and d-Pro12 could become clearer when investigating the temperature-dependent behaviour of their ions; such experiments will be described in the next chapter.

3.10 Summary

In this chapter gas-phase conformational characteristics of the Trp cage miniprotein at 300K have been investigated by IMMS and HDX. A rudimentary correlation of experimental observables to non-covalent structure has been made by means of conformational sampling by simulated annealing and MD simulations. All methods agree in one point: Trp cage ions undergo a very significant structural collapse in the absence of solvent. The compact structure of gas-phase Trp cage ions produced by nano-ESI and sprayed from near-neutral pH is reduced by the addition of extra charging protons or capping groups, but is not appreciably affected by Arg-Lys substitution or even inclusion of a d-amino acid; accordingly, simulations indicate that charge location exerts a major effect on the resultant conformation in most cases. Simulated annealing structures of Trp cage adopt a U-like architecture, occasionally ornamented by additional turns within the peptide stretches flanking the central region. Although solution-like secondary structures are exhibited in a subset of simulated annealing structures and in all configurations resulting from MD simulations at 300K, their packing is likely to be altered *in vacuo*, yielding the compact structures observed experimentally, stabilised by strong electrostatics instead of van der Waals interactions.

Isolated side-chain proton affinity has been found to correlate qualitatively with early hydrogen-deuterium exchange rates and the dissociation energy of capped $[M+2H]^{2+}$

ions (but not $[M+3H]^{3+}$ ions) as well as helical propensity of both free- and capped-termini peptides, and tertiary structural organisation of free-termini peptides (but not capped-termini peptides) as inferred from resulting CD spectra. CID and HDX data agree with predictions from the mobile proton model; furthermore results indicate that initial hydrogen-deuterium exchange events occur proximally to charged (protonated) sites. In solution (but also in the absence of solvent), the presence of more basic residues at positions 8 and 16 stabilise the salt-bridge formed between these sites and Asp9 (however zwitterions are believed to be metastable *in vacuo*[30, 45]) and also encourage helix-formation upstream of residue 8 by formation of hydrogen bonds with the exposed backbone carbonyl groups at the C-terminal end of the helix. CD experiments also show that, as expected, introduction of d-Pro12 destabilises the Trp cage fold in solution.

Removal of solvent effects significant modifications on the conformation of Trp cage ions, which undergo a structural collapse. How stable are the resulting compact configurations and what interactions are crucial for stabilising the gas-phase structure? These issues are addresses in Chapter 4 by IMMS experiments and MD simulations in vacuo at incrementally increasing temperatures.

3.11 References

1. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **1996,** *14* (1), 33-&.
2. Neidigh, J.; Fesinmeyer, R.; Andersen, N., Designing a 20-residue protein. *NAT STRUCT BIOL* **2002,** *9* (6), 425-430.
3. Neidigh, J.; Fesinmeyer, R.; Prickett, K.; Andersen, N., Exendin-4 and glucagon-like-peptide-1: NMR structural comparisons in the solution and micelle-associated states. *BIOCHEMISTRY-US* **2001,** *40* (44), 13188-13200.
4. Zhou, R., Trp-cage: Folding free energy landscape in explicit water. *P NATL ACAD SCI USA* **2003,** *100* (23), 13280-13285.
5. Qiu, L.; Pabit, S.; Roitberg, A.; Hagen, S., Ultrafast folding of a small protein: The tryptophan cage. *BIOPHYS J* **2003,** *84* (2), 166A-166A; Qiu, L.; Pabit, S.; Roitberg, A.; Hagen, S., Smaller and faster: The 20-residue Trp-cage protein folds in 4 mu s. *J AM CHEM SOC* **2002,** *124* (44), 12952-12953.
6. Pitera, J.; Swope, W., Understanding folding and design: Replica-exchange simulations of "Trp-cage" fly miniproteins. *P NATL ACAD SCI USA* **2003,** *100* (13), 7587-7592.
7. Ahmed, Z.; Beta, I.; Mikhonin, A.; Asher, S., UV-resonance Raman thermal unfolding study of Trp-cage shows that it is not a simple two-state miniprotein. *J AM CHEM SOC* **2005,** *127* (31), 10943-10950.
8. Schug, A.; Wenzel, W.; Hansmann, U., Energy landscape paving simulations of the trp-cage protein. *J CHEM PHYS* **2005,** *122* (19), -.
9. Simmerling, C.; Strockbine, B.; Roitberg, A., All-atom structure prediction and folding simulations of a stable protein. *J AM CHEM SOC* **2002,** *124* (38), 11258-11259.

10. Schug, A.; Wenzel, W., All-atom folding of the trp-cage protein with an adpative parallel tempering method. *EUROPHYS LETT* **2004,** *67* (2), 307-313.
11. Schug, A.; Herges, T.; Wenzel, W., Reproducible protein folding with the stochastic tunneling method. *PHYS REV LETT* **2003,** *91* (15), -; Carnevali, P.; Toth, G.; Toubassi, G.; Meshkat, S., Fast protein structure prediction using Monte Carlo simulations with modal moves. *J AM CHEM SOC* **2003,** *125* (47), 14244-14245.
12. Chowdhury, S.; Lee, M.; Xiong, G.; Duan, Y., Ab initio folding simulation of the Trp-cage mini-protein approaches NMR resolution. *J MOL BIOL* **2003,** *327* (3), 711-717.
13. Snow, C.; Zagrovic, B.; Pande, V., The Trp cage: Folding kinetics and unfolded state topology via molecular dynamics simulations. *J AM CHEM SOC* **2002,** *124* (49), 14548-14549.
14. Iavarone, A.; Patriksson, A.; van der Spoel, D.; Parks, J., Fluorescence Probe of Trp-cage Protein Conformation in Solution and in Gas Phase
(article in press) **2006**.
15. Adams, C.; Kjeldsen, F.; Patriksson, A.; van der Spoel, D.; Graslund, A.; Papadopoulos, E.; Zubarev, R., Probing solution- and gas-phase structures of Trp-cage cations by chiral substitution and spectroscopic techniques. *INT J MASS SPECTROM* **2006,** *253* (3), 263-273.
16. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P., A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry* **2003,** *24* (16), 1999-2012.
17. Lin, M. X.; Ahmed, Z.; Taormina, C. R.; Somayajula, K. V., A quadrupole/time-of-flight mass spectrometry study of Trp-cage's conformation. *Journal of the American Society for Mass Spectrometry* **2007,** *18* (2), 195-200.
18. Adams, C.; Kjeldsen, F.; Zubarev, R.; Budnik, B.; Haselmann, K., Electron capture dissociation distinguishes a single D-amino acid in a protein and probes the tertiary structure. *J AM SOC MASS SPECTR* **2004,** *15* (7), 1087-1098.
19. Patriksson, A.; Adams, C.; Kjeldsen, F.; Raber, J.; van der Spoel, D.; Zubarev, R., Prediction of N-C-alpha bond cleavage frequencies in electron capture dissociation of Trp-cage dications by force-field molecular dynamics simulations. *INT J MASS SPECTROM* **2006,** *248* (3), 124-135.
20. Iavarone, A.; Parks, J., Conformational change in unsolvated Trp-cage protein probed by fluorescence. *J AM CHEM SOC* **2005,** *127* (24), 8606-8607.
21. Patriksson, A.; Marklund, E.; van der Spoel, D., Protein structures under electrospray conditions. *Biochemistry* **2007,** *46* (4), 933-945.
22. Chatterjee, C.; Gerig, J. T., Interactions of hexafluoro-2-propanol with the Trp-cage peptide. *Biochemistry* **2006,** *45* (49), 14665-14674; Chatterjee, C.; Gerig, J. T., Interactions of trifluroethanol with the Trp-cage peptide. *Biopolymers* **2007,** *87*, 115-123; Neuman, R. C.; Gerig, J. T., Solvent interactions with the Trp-cage peptide in 35% ethanol-water. *Biopolymers* **2008,** *89* (10), 862-872.
23. Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H., Designing a 20-residue protein. *Nature Structural Biology* **2002,** *9* (6), 425-430.
24. Adams, C. M.; Kjeldsen, F.; Patriksson, A.; van der Spoel, D.; Graslund, A.; Papadopoulos, E.; Zubarev, R. A., Probing solution- and gas-phase structures of Trp-cage cations by chiral substitution and spectroscopic techniques. *International Journal of Mass Spectrometry* **2006,** *253* (3), 263-273.
25. Patriksson, A.; Adams, C.; Kjeldsen, F.; Raber, J.; van der Spoel, D.; Zubarev, R. A., Prediction of N-C-alpha bond cleavage frequencies in electron capture dissociation of Trp-cage dications by force-field molecular dynamics simulations. *International Journal of Mass Spectrometry* **2006,** *248* (3), 124-135.
26. Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F., Structural information from ion mobility measurements: Effects of the long-range potential. *Journal of Physical Chemistry* **1996,** *100* (40), 16082-16086.
27. Adams, C. M.; Kjeldsen, F.; Zubarev, R. A.; Budnik, B. A.; Haselmann, K. F., Electron capture dissociation distinguishes a single D-amino acid in a protein and probes the tertiary structure. *Journal of the American Society for Mass Spectrometry* **2004,** *15* (7), 1087-1098.

28. Juraszek, J.; Bolhuis, P. G., Sampling the multiple folding mechanisms of Trp-cage in explicit solvent. *Proceedings of the National Academy of Sciences of the United States of America* **2006,** *103* (43), 15859-15864.
29. Hudaky, P.; Straner, P.; Farkas, V.; Varadi, G.; Toth, G.; Perczel, A., Cooperation between a salt bridge and the hydrophobic core triggers fold stabilization in a Trp-cage miniprotein. *Biochemistry* **2008,** *47* (3), 1007-1016.
30. Price, W. D.; Jockusch, R. A.; Williams, E. R., Is arginine a zwitterion in the gas phase? *Journal of the American Chemical Society* **1997,** *119* (49), 11988-11989.
31. Schnier, P. D.; Price, W. D.; Jockusch, R. A.; Williams, E. R., Blackbody infrared radiative dissociation of bradykinin and its analogues: Energetics, dynamics, and evidence for salt-bridge structures in the gas phase. *Journal of the American Chemical Society* **1996,** *118* (30), 7178-7189.
32. Hudgins, R. R.; Ratner, M. A.; Jarrold, M. F., Design of helices that are stable in vacuo. *Journal of the American Chemical Society* **1998,** *120* (49), 12974-12975.
33. Harrison, A. G., The gas-phase basicities and proton affinities of amino acids and peptides. *Mass Spectrometry Reviews* **1997,** *16* (4), 201-217.
34. Kelly, S. M.; Price, N. C., The application of circular dichroism to studies of protein folding and unfolding. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology* **1997,** *1338* (2), 161-185.
35. Dongre, A. R.; Jones, J. L.; Somogyi, A.; Wysocki, V. H., Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: Evidence for the mobile proton model. *Journal of the American Chemical Society* **1996,** *118* (35), 8365-8374.
36. Campbell, S.; Rodgers, M. T.; Marzluff, E. M.; Beauchamp, J. L., Deuterium exchange reactions as a probe of biomolecule structure. Fundamental studies of cas phase H/D exchange reactions of protonated glycine oligomers with D2O, CD3OD, CD3CO2D, and ND3. *Journal of the American Chemical Society* **1995,** *117* (51), 12840-12854.
37. Wagner, D. S.; Anderegg, R. J., CONFORMATION OR CYTOCHROME-C STUDIED BY DEUTERIUM-EXCHANGE ELECTROSPRAY-IONIZATION MASS-SPECTROMETRY. *Analytical Chemistry* **1994,** *66* (5), 706-711.
38. Florance, H. V. Exploring protein conformations with mass spectrometry. Thesis (Ph D), University of Edinburgh, 2008.
39. Iavarone, A. T.; Parks, J. H., Conformational change in unsolvated Trp-cage protein probed by fluorescence. *Journal of the American Chemical Society* **2005,** *127* (24), 8606-8607.
40. Wyttenbach, T.; Paizs, B.; Barran, P.; Breci, L.; Liu, D. F.; Suhai, S.; Wysocki, V. H.; Bowers, M. T., The effect of the initial water of hydration on the energetics, structures, and H/D exchange mechanism of a family of pentapeptides: An experimental and theoretical study. *Journal of the American Chemical Society* **2003,** *125* (45), 13768-13775.
41. Ewing, N. P.; Pallante, G. A.; Zhang, X.; Cassady, C. J., Gas-phase basicities for ions from bradykinin and its des-arginine analogues. *Journal of Mass Spectrometry* **2001,** *36* (8), 875-881.
42. Fernandez, A.; Colubri, A.; Berry, R. S., Topologies to geometries in protein folding: Hierarchical and nonhierarchical scenarios. *Journal of Chemical Physics* **2001,** *114* (13), 5871-5887.
43. Dobo, A.; Kaltashov, I. A., Detection of multiple protein conformational ensembles in solution via deconvolution of charge-state distributions in ESI MS. *Analytical Chemistry* **2001,** *73* (20), 4763-4773.
44. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., ELECTROSPRAY IONIZATION-PRINCIPLES AND PRACTICE. *Mass Spectrometry Reviews* **1990,** *9* (1), 37-70.
45. Jarrold, M. F., Peptides and proteins in the vapor phase. *Annual Review of Physical Chemistry* **2000,** *51*, 179-207.

4

Thermal Unfolding of Trp Cage *in vacuo*

4.1 Abstract

Structural traits of the Trp cage miniprotein, at room temperature were investigated in Chapter 3. Therein it is established that even those peptide constructs which do not fold stably in solution adopt configurations more compact than the solution fold in the gas phase. Peptides with protected termini tend to be more extended relative to constructs with free termini. Still, there is no clear correlation between the trends observed for Arg- and Lys-substituted peptides observed by CID, HDX or IMMS. In order to investigate the thermodynamic stability of the peptide's structure, high-temperature measurements and simulations were performed and are reported in this chapter. Thermal unfolding of Trp cage in solution was explored by CD. The temperature dependence of the peptides' configuration *in vacuo* has been investigated by ion mobility measurements; the thermodynamic stability of the Trp cage fold was also followed by means of MD simulations.

CD experiments show that Trp cage behaves, at least macroscopically, in a way consistent with that of a co-operative two-state folder. IMMS and MD data confirm that the gas-phase conformational stability is dominated by electrostatics; additional charge or incorporation of terminal protecting groups yield dramatic effects on the overall architecture of Trp cage constructs. Different constructs exhibit considerable similarities

at highest temperatures in the studied range, characteristic of fast and relatively large-scale conformational fluctuations occurring at such temperatures. Computational work unveils the importance of different intramolecular interactions for stabilising a fold in the gas phase, a result that has implications for the study of solution-like non-covalent structure in biomolecular ions produced by electrospray ionisation.

4.2 Introduction

Both computational and experimental efforts have been invested previously towards understanding the structural behaviour of Trp cage at elevated temperatures in the presence and absence of water. As explained previously (section 3.2) Trp cage has served as a model system for testing algorithms that can be of value in the study of protein folding, and features in numerous simulations in the computational chemistry arena. The first such studies employed replica-exchange for temperature-dependent conformational sampling[1, 2]. Pitera and Swope[1] employed the Cornell *et al.* force field[3] with a generalised Born solvent whereas Zhou[2] used the OPLS-AA force field and the SPC explicit water model. It is obvious in both studies that current parameterisations overestimated the thermal stability of the Trp cage, where a drop in the population of native-like traits was observed to start at temperatures around (or even higher than) 400K, far above the experimentally determined "melting" temperature of ~315K[1, 2]. Nevertheless, a very good qualitative agreement between theory and experiment could be drawn.

Zhou's REM simulations[2] allowed the construction of two-dimensional free energy maps for the Trp cage at different temperatures. The free energy landscape appeared to be more 'rugged' at lower temperatures, becoming increasingly smooth and funnel-like as the temperature was elevated but remaining below the 'melting' temperature. This finding supports the presence of partially folded structures at which the system can be trapped at low temperatures. The author noted the presence of a transient folding intermediate containing two hydrophobic cores *separated* by the D9-R16 salt bridge[2] consistent with the hypothesis that residual, "native-like" contacts are present in the "unfolded" peptide[1]. The barrier between this conformer and the "native" state must be too small to delay the folding reaction.

More recently, Hudáky *et al.*[4] investigated Trp cage thermodynamics using CD and NMR measurements. Results on several constructs, including the D9N back-mutant and a glycosylated variant, were reported. Analysis of far-UV CD spectra allowed the determination of different secondary structural elements. Furthermore, the authors

embarked on a further protein design effort and concluded that a D9E substitution improves the thermodynamic stability of the fold; presumably this effect can be attributed to a 'release' from a less sterically favourable conformation of K8 and R16 side-chains participating in the K8-D9-R16 salt bridge.

FRET experiments carried out by Iavarone *et al.*[5-8] in solution also confirm Trp cage as a two-state folding system. On the other hand the D9N construct does not form a stable structure. The authors carried out FRET measurements *in vacuo* as well; the results of this study are reviewed in section 3.2. Yet the main finding was that in the $[M+3H]^{3+}$ charge state Trp6 indole becomes exposed to collisions with the BODIPY-TMR chromophore at lower temperatures than in the $[M+2H]^{2+}$ charge state; furthermore, ions produced for less well-folding peptides, such as D9N, were found to be only slightly less thermally resistant. In both cases, the midpoint of the 'unfolding transition' inferred occurred at temperatures much higher than the solution melting temperature of ca. 310K.

REM simulations of Trp cage *in vacuo* have also been performed[8-10] to support both the fluorescence data described by Iavarone *et al.* [8] and Adams *et al.* [9, 11]. In these studies it is found that equilibrium structures of solvent-free Trp cage ions bear significant similarities to the solution structure as far as secondary structure is concerned; however tertiary contacts relying on van der Waals interactions are lost. Analysis of these trajectories led the authors to stress the importance of hydrogen-bonded structures in stabilising macromolecular conformations in the absence of solvent[8-10].

In this chapter Trp cage is shown to conform macroscopically to a two-state folding model using CD. IMMS and MD data ensue, in experiments performed in order to assess how the intrinsic stability of the peptide is affected by the following:

1. charge and charge location;
2. terminal capping groups;
3. K→R and R→K substitutions and
4. incorporation of d-Pro12.

4.3 Thermal unfolding of Trp cage in Solution as Probed by CD

Absorption bands in the CD spectra can be attributed as characteristic of secondary structure – for example helicity, yielding a band around 220nm – or tertiary contacts – for example the band around 282nm, attributable to dichroism of Trp6. Monitoring the temperature-dependent evolution of these spectral features enables thermodynamic analysis of their corresponding structural elements. Data were collected on a Jasco J-810 spectropolarimeter; sample temperatures were controlled by means of a peltier temperature regulator. Parameters for these measurements are listed in Table 4.1

Table 4.1
Parameters for temperature-dependent CD measurements

Heating Rate	Data pitch	Temperature Range	Wavelengths studied
0.5K/min	2K	288.15-358.15K (15-85°C)	220nm and 282nm

Temperature-dependence experiments yielded broad sigmoidal curves which where fitted and analysed using a simple two-state model. The resulting van ’t Hoff plots can be seen in Figure 4.1. Estimates of folding enthalpies and entropies for the two wavelengths are displayed in Table 4.2.

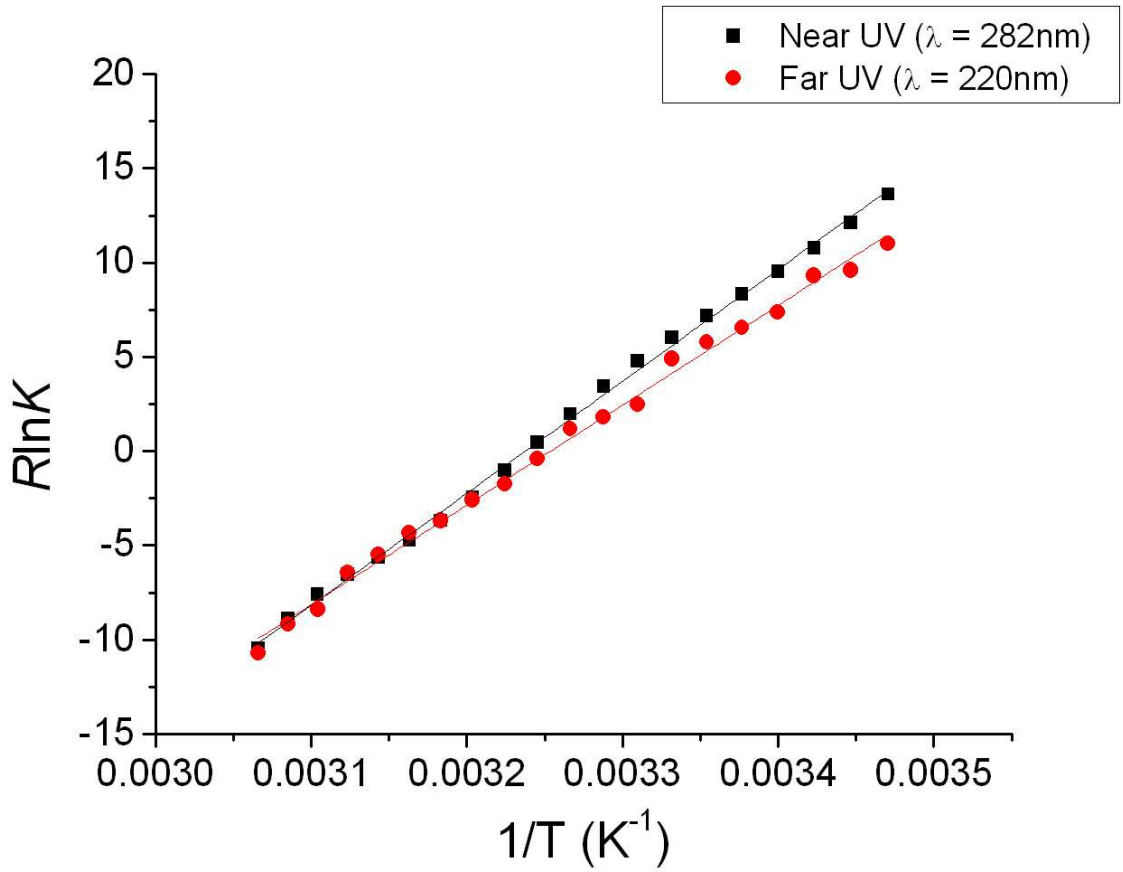


Figure 4.1 van 't Hoff plot of Trp cage tc5b (WT) unfolding derived from CD data collected at 220nm (red circles) and 282nm (black squares).

Table 4.2
Thermodynamic data for Trp cage determined by CD

λ / nm	Assignment	ΔH / kJ·mol^{-1}	ΔS / kJ·mol^{-1}·K^{-1}	T_m / K
220	Helix	−52.98 ± 1.42	−0.1724 ± 0.0046	307.3 ± 16.5
282	Hydrophobic core	−59.35 ± 0.71	−0.1921 ± 0.0023	308.9 ± 7.4

Data obtained clearly indicate unfolding of the peptide at elevated temperatures; as a matter of fact the 'melting temperature' (temperature at which folded and unfolded ensembles are equally populated) is larger than room temperature only by ~10K. The poor stability of the fold at elevated temperatures in aqueous solution (10mM ammonium

acetate) may have implications for ions produced by nano-ESI. All mass spectrometers used contain heated sources to assist with desolvation of analyte ions; however if the temperature in the analyte droplets exceeds by far room temperature then there is danger that folded peptides are not transferred in the gas phase. On the other hand, the temperature of the analytes tends to be lowered by solvent evaporation from the droplets[12]. Thermal unfolding curves obtained at 220nm and 282nm are very similar, likewise yielding similar thermodynamic quantities. Such similarity in the thermodynamic behaviour of secondary and tertiary contacts indicates the co-operativity of the folding process, a finding that is also supported by the fact that the d-Pro12 construct lacks both stable 3° *and* 2° structure. Nevertheless certain differences do manifest themselves. Native-like hydrophobic contacts involving W6 (and possibly Y3) are characterised by a smaller enthalpy of formation but also with smaller entropy compared to those obtained for helices, indicating that folding is hydrophobically driven. If the difference in T_m of the two curves is to be formally interpreted is suggests that in a small temperature interval (between 307.3 and 308.9K) the peptide adopts a state that still contains a hydrophobic core but not much secondary structure, a scenario that has been hinted to by both computational[13] and experimental[14] studies. Yet, the difference in T_m of helices and hydrophobic contacts involving W6 is too small compared to estimated error to be deemed significant.

Overall, thermodynamic measurements of Trp cage using CD show the co-operative nature of the peptide's fold, with secondary and tertiary elements forming concurrently and helical regions even requiring the presence of a hydrophobic centre in relation to which they are organised. On the other hand, gas-phase architectures relying on either H-bonding (e.g. secondary structure) or van der Waals interactions (the majority of the tertiary contacts in Trp cage) are likely to "melt" at different energies: thus, their temperature dependence is expected to be uncorrelated.

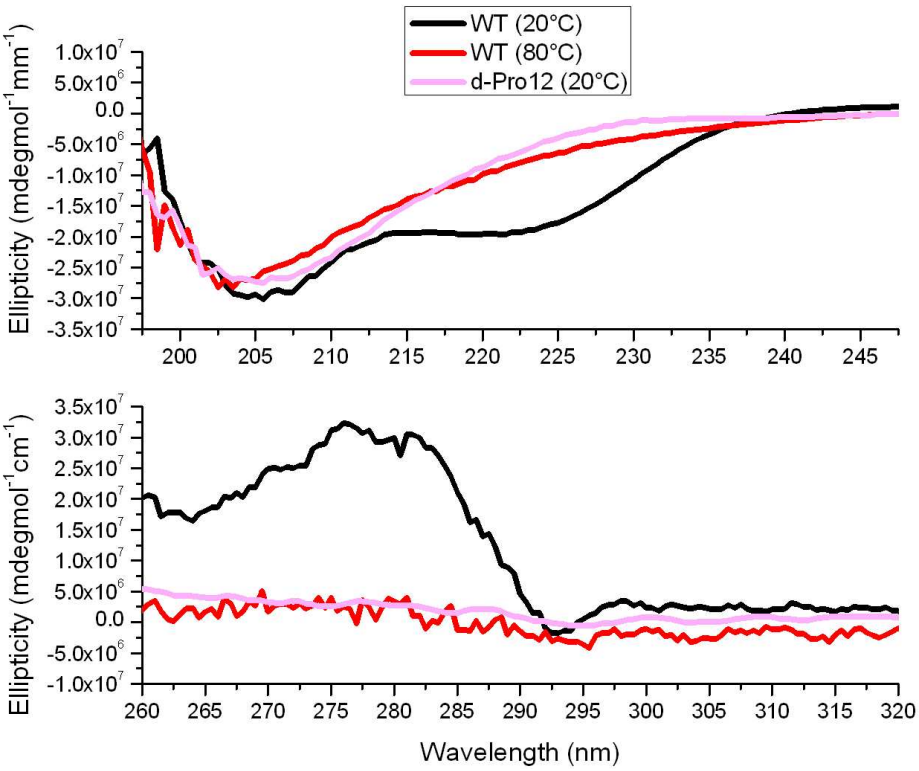


Figure 4.2 Far-UV (top) and Near-UV (bottom) spectra of WT Trp cage (free termini) at 25℃ (287.15K, black lines) and 80℃ (353.15K, red lines). Spectra of d-Pro12 (pink lines) at 25℃ are also shown for comparison. Note that minor differences are detected between the dichroic properties of WT at high temperature and d-Pro12, suggesting Trp cage preserving some residual structure at high temperatures in 10mM aqueous NH_4CH_3COO.

4.4 Effect of Charge and Terminal Protecting Groups on Trp cage Structure

4.4.1 Ion Mobility of WT Trp Cage Ions at Elevated Temperatures

Coulombic repulsion upon addition of charge is a well-established trait of gas-phase peptide ions, which has been readily observable by IMMS[15]; it is also a well documented phenomenon in CID, IMS experiments and molecular mechanics simulations described in the previous chapter. Here charge is indeed found to be influential to peptide structure, altering temperature-dependent structural fluctuations of Trp cage ions, as seen in Figure 4.3. The method for data collection is described in Chapter 2; temperature-dependent

IMS experiments were achievable by use of resistive heating. The temperature was monitored by means of thermocouples located on the cell body and the end cap; during experiments temperature fluctuations were small and did not exceed 3K. Experimental uncertainty of these measurements is large. There are a number of reasons that can explain for this event. The most likely cause is the variation in instrument parameters. Some parameters, such as source pressure, were very hard to reproduce and had to be adjusted in order to optimise ion transmission. As the preservation of non-covalent aggregates is influenced by source pressure[16] the resulting cross sections may vary accordingly, due to the complex effects of in-source non-specific aggregate dissociation.

Over the studied temperature range (300-550K) the $[M+3H]^{3+}$ ions start with and retain a more extended conformation than $[M+2H]^{2+}$ ions. In addition thermally-induced conformational changes occur at different rates between the two charge states. In peptides with unprotected termini, $[M+2H]^{2+}$ ions adopt very compact configurations (more compact than the solution Trp cage fold by *ca.* 60Å^2) that unfold only at temperatures above 450K. Conversely, $[M+3H]^{3+}$ ions become gradually more extended at temperatures higher than 350K, after a possible modest structural collapse between 300 and 350K. Notably, transitions to observably more extended structures take place at temperatures much higher than the peptide's T_m in solution as determined by CD (Table 2.2). This observation can lead to two possible, not mutually exclusive, conclusions:

1. Conformational free energy landscapes of peptide ions become more 'glass-like' in the absence of solvent; as interactions with solvent molecules are not possible, looser structures (conformational transition states) are less stable. This effect is intensified by Coulombic repulsion.
2. 'Non-native', 'unfolded' structures may be very compact, as a matter of fact more compact than the 'native' state. The necessity of self-solvation in the gas phase stabilises such structures. Given the great extent of structural collapse observed, this conclusion seems to be more likely.

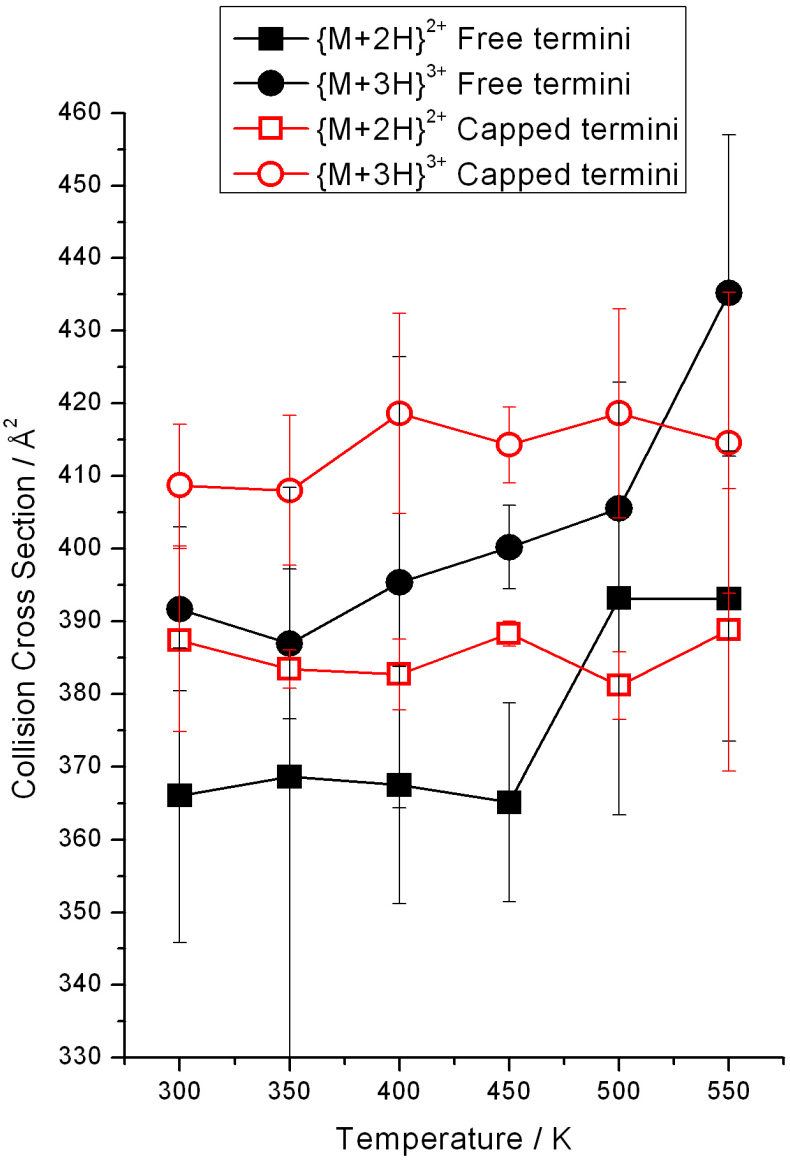


Figure 4.3 Collision cross sections of protonated species of Trp cage (WT) as a function of temperature. Filled symbols: constructs with free termini; open symbols: constructs with protected termini. Squares: $[M+2H]^{2+}$ ions; circles: $[M+3H]^{3+}$ ions.

Trp cage ions with protected termini display significant differences to their unprotected analogues. These differences are two-fold; firstly, both charge states of capped peptides are more extended than the corresponding ions of free-termini peptides; secondly, their fluctuations with temperature differ as well. $[M+2H]^{2+}$ ions become slightly smaller with

temperature (a decrease in cross section is expected because the interaction between ions and neutrals becomes less pertinent as the temperature is increased), with measurements taken at 450 and 550K being possible exceptions; let it be said that this decrease is within error and therefore questionable. $[M+3H]^{3+}$ charge states display a relatively small increase in cross section between 350K and 450K, above which temperatures they do not change. Cross sections for protected peptides do not exceed that estimated for the solution fold (420Å^2), implying that a structural collapse has taken place upon desolvation, yet it is not nearly as dramatic as the one observed for free-termini constructs.

The main difference between Trp cage constructs with and without protected termini is the proton affinity of the termini, especially the N-terminus. Addition of an N-terminal acetyl group effects a significant decrease in gas-phase basicity of the N-terminus; thus other protonation states become more populated. This effect is quite significant for $[M+3H]^{3+}$ ions. Capping groups also affect the strength of hydrogen bonds involving one or both peptide termini, which may well be the cause responsible the for observed difference between capped and free-termini $[M+2H]^{2+}$ cross sections.

Overall, the change observed upon addition of terminal capping groups can be attributed to 1) perseverance of native-like characteristics in the gas phase (as is the case in solution, verified by CD, as seen in the previous chapter) or 2) destabilisation of compact gas-phase structures adopted by the uncapped peptide construct by perturbation of electrostatic interactions involving the peptide termini. MD simulations can be used to gain insight into the dependence of folded state stability on charge location and the presence of terminal protecting groups *in vacuo*.

4.4.2 Overview of Gas-phase MD Simulations on WT Trp Cage

Constant-temperature MD were performed at 50K increments starting at 300K the highest temperature simulated being 600K; the NMR structure served as a starting geometry, and was manipulated for the simulation as described in Chapter 2. Each constant-temperature run was 10ns long, using a 0.5fs time-step. Thus, each construct

was simulated for 70ns ($1.4 \cdot 10^8$ MD iterations), excluding preparatory heating and equilibration MD. Constructs simulated are listed in Tables 3.3 and 3.8; important features trajectories generated from simulations of WT Trp cage are displayed in Figures 4.4-4.11.

It is indeed worth noting that trans-conformational events do not necessarily result into more extended structures. Figures 4.4 and 4.5, displaying collision cross sections (calculated with *sigma*) and the radius of gyration of simulated systems, contain examples of, often significant, structural collapse taking place at low and intermediate temperatures in the studied range. Such changes are observed in all simulations, often temporally correlated to a departure from the solution Trp cage fold. Furthermore, in systems where such a departure has occurred at the heating equilibration stages, such as WT2a & 2b, the resulting configurations are more compact than the NMR structure. Structural changes often proceed via extended structures, but lead to more compact configurations. Qualitatively this effect agrees with experiment, as all cross sections measured at temperatures below 550K are smaller than that estimated for the NMR structure. Comparison of Figures 4.4 (cross sections) and 4.5 (radii of gyration) reveals the expected decrease of cross section with temperature, attributable to the energetics of He-analyte collisions (attractive forces between two colliding atoms, arising primarily due to dispersion, becaome weaker compared to k_BT as T increases); however Figure 4.5 clearly indicates the decrease in cross section observed also results from conformational changes.

Simulations of WT Trp cage clearly imply that the stability of the solution conformation in the absence of water is highly dependent on charge. As seen in Figure 4.6, the temperature at which the backbone topology has changed significantly from that of the solution fold (reaching a heavy atom backbone RMSD > 3Å) differs depending on charge assignment. WT2a and 2b have undergone such transitions during the initial heating stages. Forcing zwitterion formation in the 2+ charge states (WT2z and 2z-b) greatly stabilised the structure; major conformational changes for these systems occur at 500 and 550K, respectively. Surprisingly, the solution fold appears to be more stable in 3+ ions

with free termini, compared to non-zwitterionic 2+ ions (with unprotected termini). WT3a does not change its backbone architecture until a simulation temperature of 450K; WT3b displays a considerable modification at 300K but retains essential features of the fold up to 450-500K; WT3c fluctuates to a greater extent and finally ‘unfolds’ (with respect to the NMR structure) at 400K.

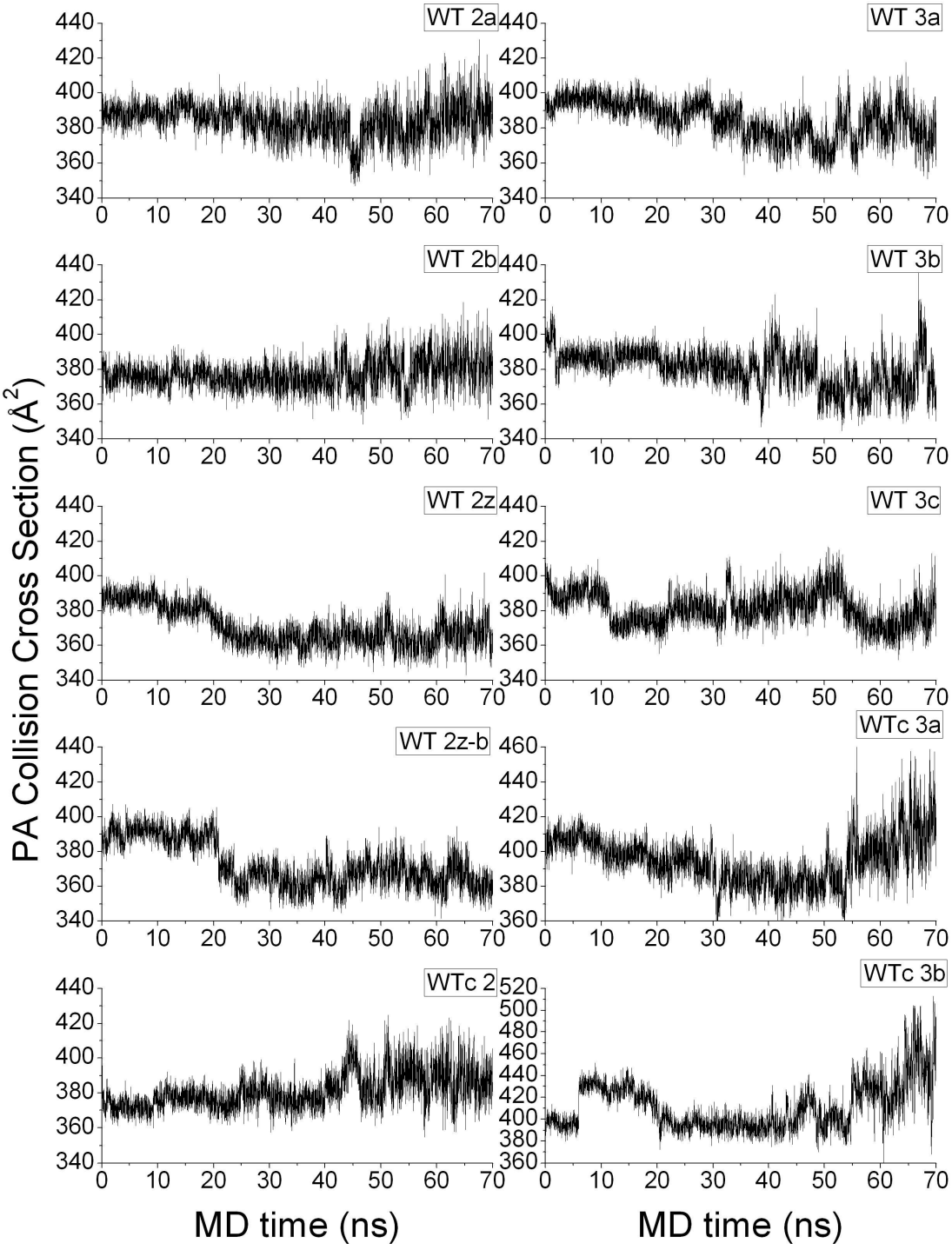


Figure 4.4
Collision cross sections for different charge variants of WT Trp cage, calculated using the projection approximation method. Cross sections were estimated every 10ps of simulation time. Simulation temperature increments by 50K every 10ns of the trajectory: 0-10ns: 300K; 10-20ns: 350K; 20-30ns: 400K; 30-40ns: 450K; 40-50ns: 500K; 50-60ns: 550K; 60-70ns: 600K. All graphs are *not* in scale.

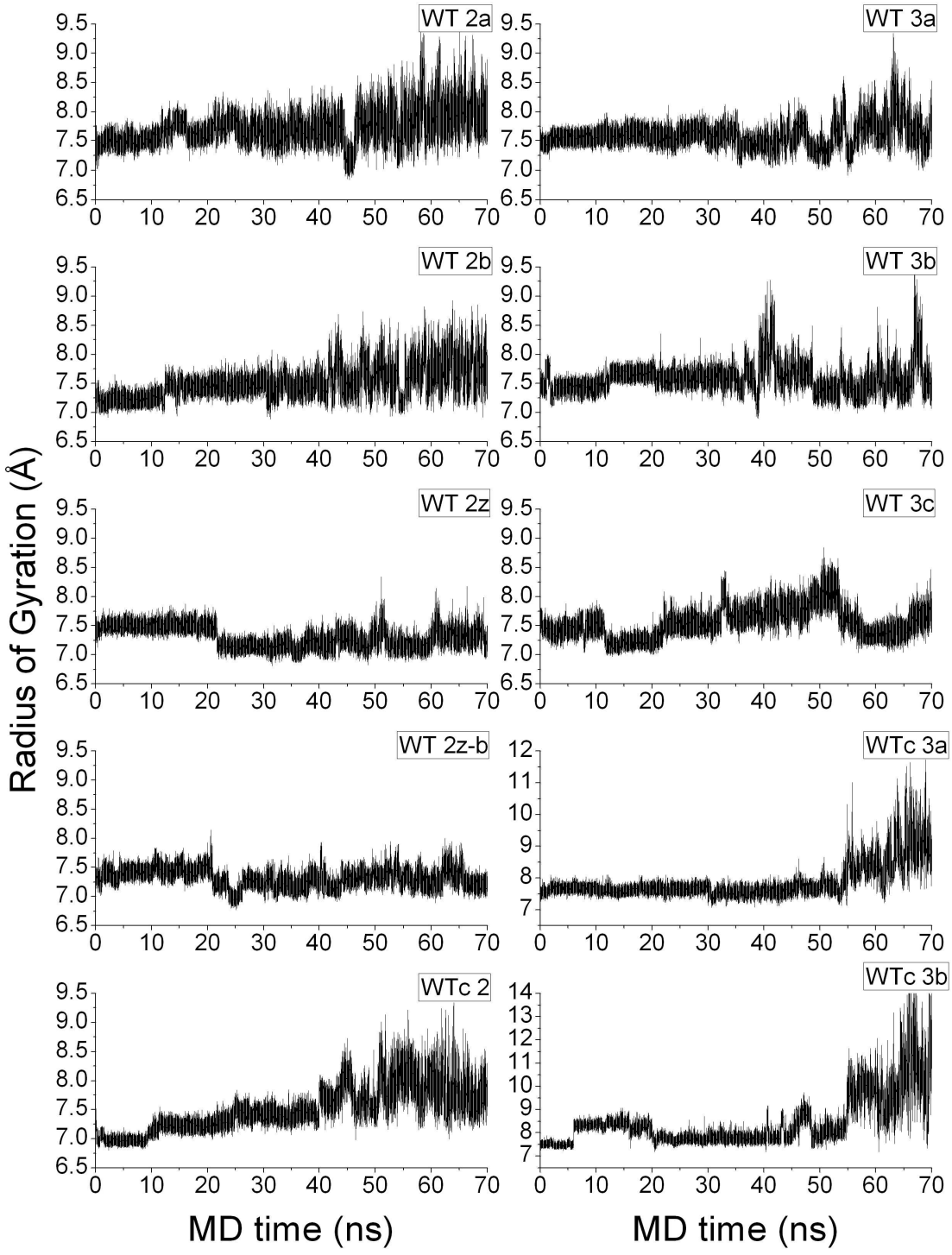


Figure 4.5
Radii of gyration for different charge variants of WT Trp cage. Simulation temperature increments by 50K every 10ns of the trajectory: 0-10ns: 300K; 10-20ns: 350K; 20-30ns: 400K; 30-40ns: 450K; 40-50ns: 500K; 50-60ns: 550K; 60-70ns: 600K. All graphs are *not* in scale.

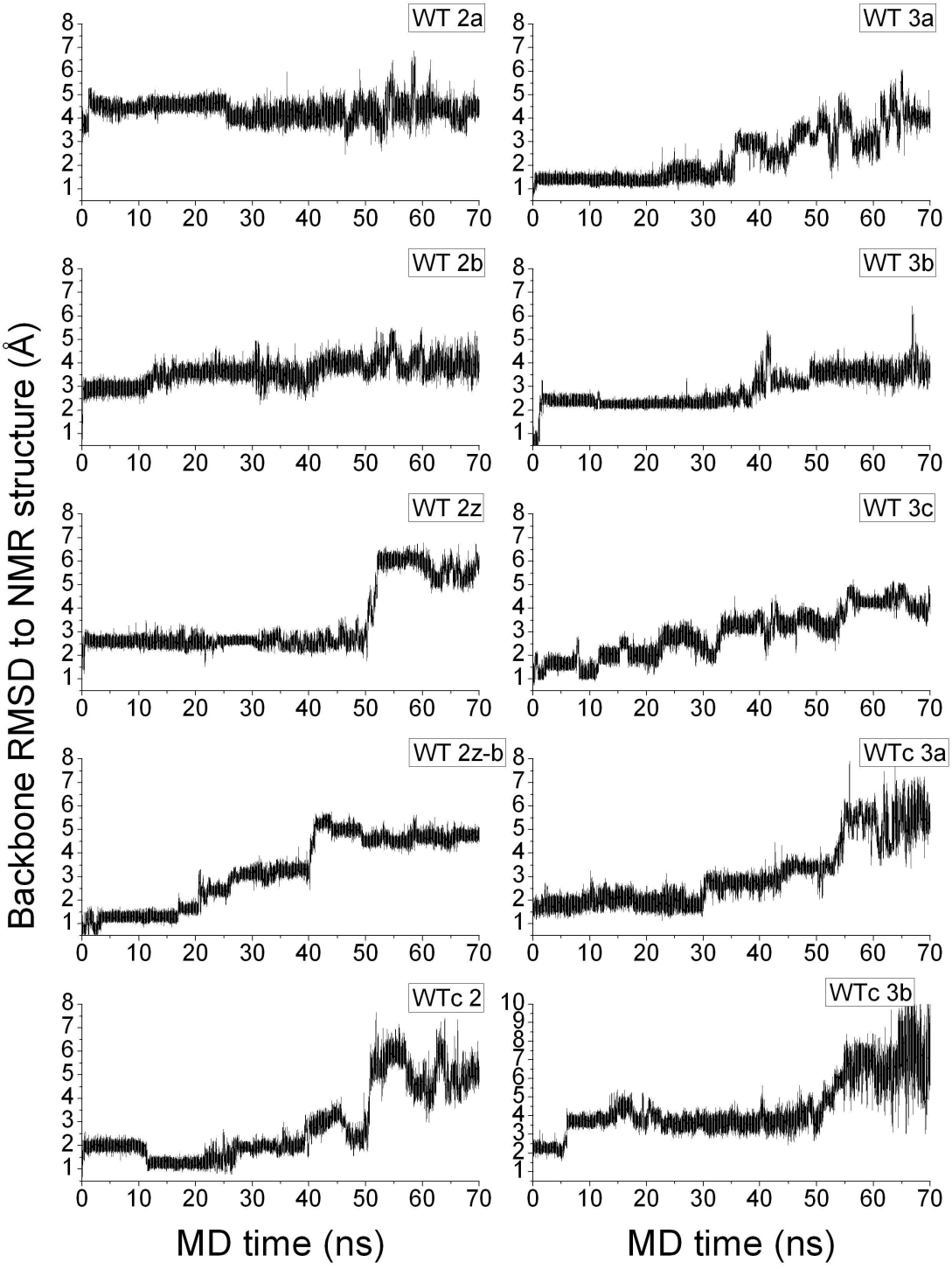


Figure 4.6
Heavy-atom backbone RMSD to the NMR structure calculate for different charge variants of WT Trp cage. Simulation temperature increments by 50K every 10ns of the trajectory: 0-10ns: 300K; 10-20ns: 350K; 20-30ns: 400K; 30-40ns: 450K; 40-50ns: 500K; 50-60ns: 550K; 60-70ns: 600K. All graphs are *not* in scale.

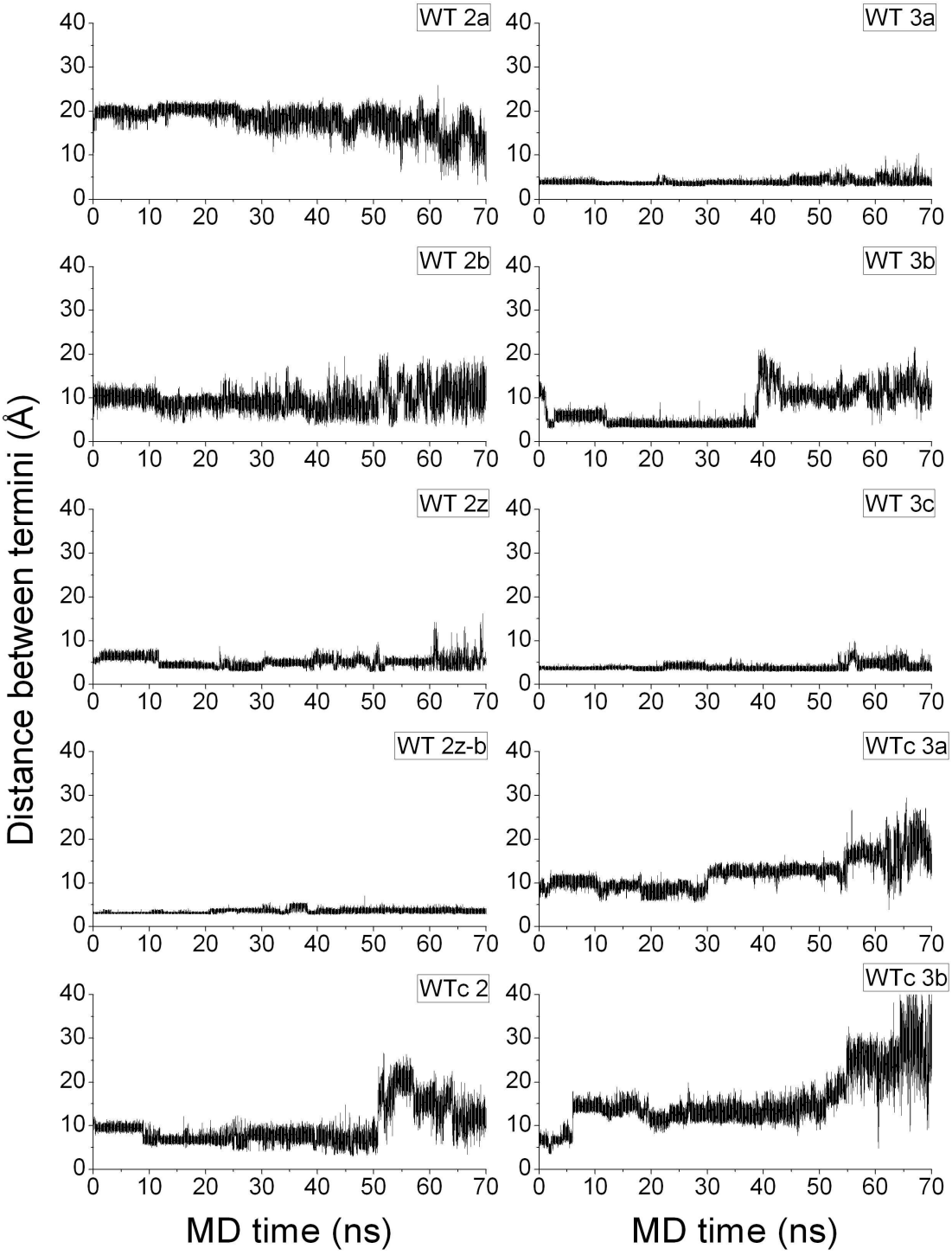


Figure 4.7
Distance between backbone nitrogen of Asn1 and backbone carbonyl carbon of Ser20, calculated for the ten charge variants of WT Trp cage. Simulation temperature increments by 50K every 10ns of the trajectory: 0-10ns: 300K; 10-20ns: 350K; 20-30ns: 400K; 30-40ns: 450K; 40-50ns: 500K; 50-60ns: 550K; 60-70ns: 600K. All graphs are in scale.

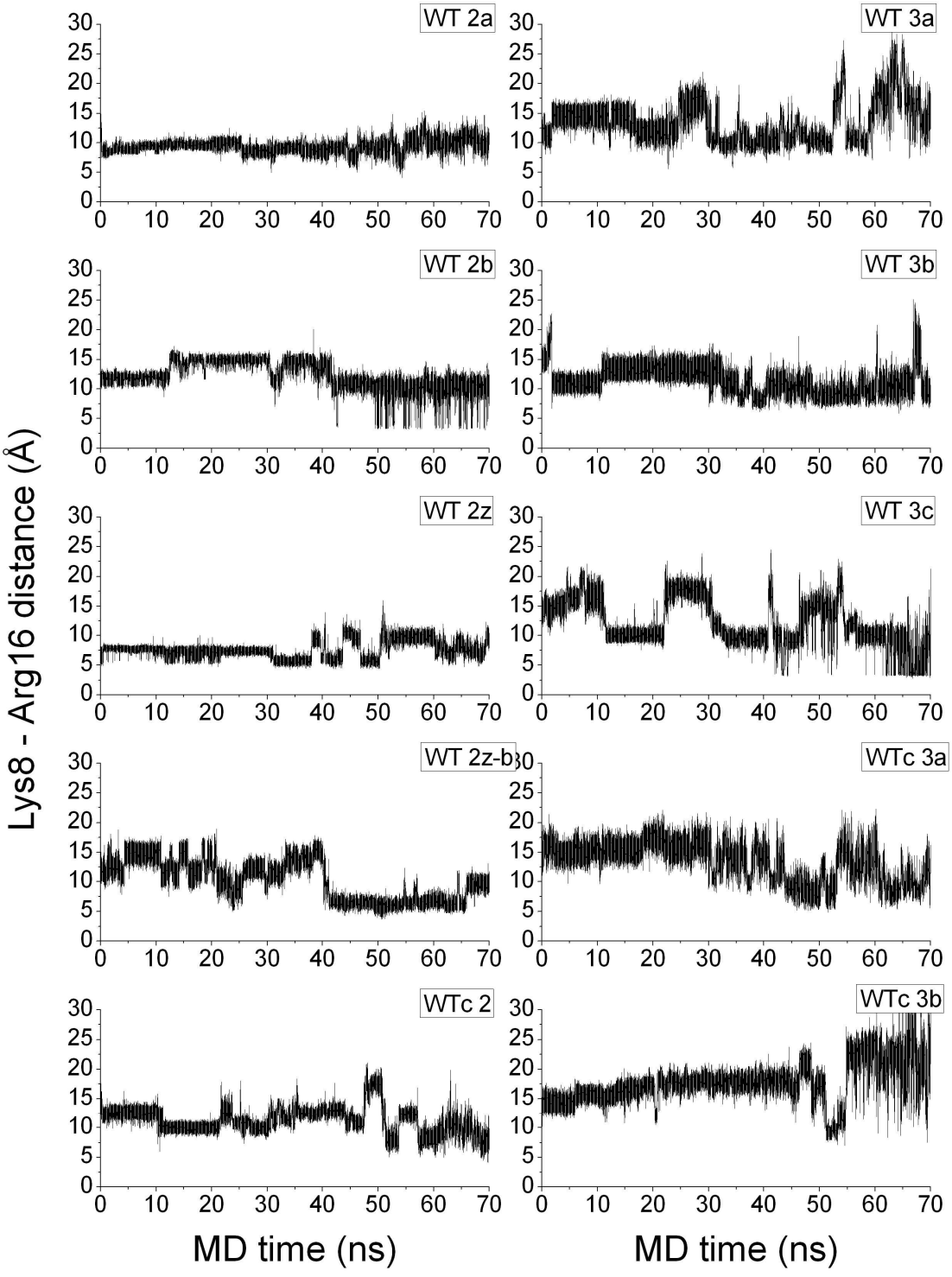


Figure 4.8
Distance between N_ζ of Lys8 and C_ζ of Arg16 for different charge variants of WT Trp cage. Simulation temperature increments by 50K every 10ns of the trajectory: 0-10ns: 300K; 10-20ns: 350K; 20-30ns: 400K; 30-40ns: 450K; 40-50ns: 500K; 50-60ns: 550K; 60-70ns: 600K. All graphs are in scale.

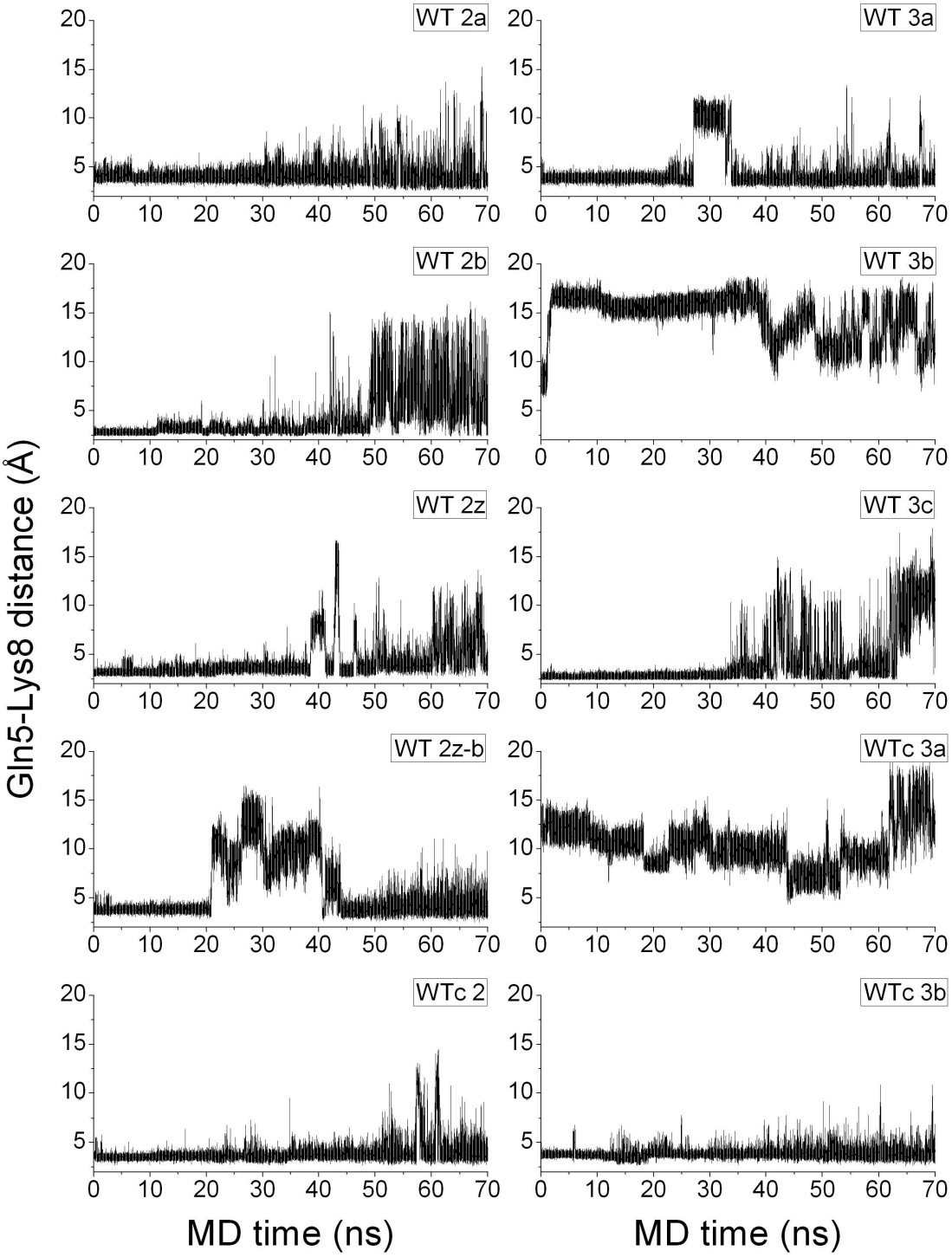


Figure 4.9
Distance between C_δ of Gln5 and N_ζ of Lys8 calculated for different charge variants of WT Trp cage. Simulation temperature increments by 50K every 10ns of the trajectory: 0-10ns: 300K; 10-20ns: 350K; 20-30ns: 400K; 30-40ns: 450K; 40-50ns: 500K; 50-60ns: 550K; 60-70ns: 600K. All graphs are in scale.

4.4.3 Molecular interactions that influence gas-phase conformational stability

Figures 4.7, 4.8 and 4.9 illustrate that gas-phase stability of Trp cage conformation depends on a number of key interactions involving (putative) protonated groups. Figure 4.8 displays the distance between N_ζ of K8 and C_ζ of R16; short distances imply that the side-chains of the two basic groups form hydrogen bonds to proximal atoms during the MD trajectory. As expected, the distance between these two atoms is short throughout the entire simulation of WT2z – indicating the persistence of the K8-D9-R16 salt bridge in this charge variant – and fluctuations in this coordinate coincide with large-scale structural changes. Nevertheless a short K8-R16 distance is also encountered in the simulation of WT2z-b, although in the latter case such interactions feature more prominently at higher temperatures ($T \geq 450K$), after the Trp cage fold has been significantly perturbed, showing that networks of interactions involving charged residues stabilise gas-phase protein structures. The distance between the terminal heavy atoms (backbone nitrogen of N1 and backbone carbonyl carbon atom of S20, shown in Figure 4.7) is more informative. H-bonding between terminal groups in peptides with uncapped termini persists not only in WT2z-b, in which an ionic interaction is formed between those groups, but also in WT2z, 3a, 3b and 3c; in short, an interaction between the termini is present in all simulations of systems in which the Trp cage fold appears to be stable or the N-terminus is protonated. As the tertiary structure of Trp cage in solution involves a U-shape overall topology, bringing the termini together, a strong interaction between the latter contributes significantly towards the stabilisation of this molecular architecture in the absence of solvent.

4.4.4 Comparison between IMMS experiments and simulations

The significance of strongly interacting termini when the N-terminus is protonated justifies the difference observed between capped and uncapped $[M+3H]^{3+}$ species, as in the former the proton affinity of the N-terminus is decreased. The combination of IMMS and MD simulations, both performed over a range of temperatures implicates the N-

terminus as a proton carrier in the 3+ charge state of Trp cage. However the discrepancy between the experimentally observed increase in cross-section with temperature for $[M+3H]^{3+}$ free-termini peptide ions and the trend observed for the three simulated $[M+3H]^{3+}$ charge isomers needs to be accounted for.

Two particularly important aspects of the experiment are not reproduced by molecular mechanics simulations. First, the simulated timescale is smaller than the experimental one by *ca*. five orders of magnitude – let us re-iterate that these simulations only aim for conformational sampling in the vicinity of the solution fold on the gas-phase free energy surface. Significant conformational changes may take place in the experimental time-scale at lower temperatures than these predicted from the simulations presented here. The second aspect is that of intramolecular proton transfer – notably no single charge assignment reproduces experimental trends; however, especially for $[M+2H]^{2+}$ charge states, WT2a & 2b trajectories agree with experiment at high temperatures. Increased rates of proton transfer at high temperatures as well as the co-existence of several charge 'isomers' produced at low temperatures by ESI are likely to account for the simulation trends of $[M+2H]^{2+}$ ions. Additionally, CID experiments (section 3.8.2) clearly indicate the importance of proton mobility in desolvated Trp cage ions: a higher rate of intramolecular proton transfer *in vacuo* for $[M+3H]^{3+}$ species is implied by the lower energies required for fragmentation and the independence of this fragmentation to subtle changes to the peptide's constituent amino acids' gas-phase proton affinity. Yet the stabilisation of the Trp cage fold achieved by strengthening the interaction between the termini upon protonation of the N-terminus should be observable if $[M+3H]^{3+}$ ions preserved the solution fold in the gas-phase, as the NMR-determined configuration occupies higher parts of cross-section space than optimal gas-phase structures (section 3.5). Thus we are led to suggest the following:

Given the modest thermal stability of WT Trp cage (section 4.2) a significant population of peptides might be unfolded before or during the ESI process. CD and charge-state distributions of electrosprayed ions of WT Trp cage compared with d-Pro12 (section 3.9) support the existence of a correlation between unfolded states of the peptide

present in solution and the relative intensity of the 3+ charge state, a relation that has also been realised elsewhere[17]. Therefore, the absence of $[M+3H]^{3+}$ ions preserving the solution conformation experimentally could be attributed to the fact that $[M+3H]^{3+}$ species originate predominantly from 'unfolded' or randomly quenched population of Trp cage molecules in solution. Early ESI-IMS-MS studies on protein systems clearly showed polypeptide unfolding taking place as a function of charge[15]. This effect is partially brought about by gas-phase coulombic effects. However, the presence of several features in protein charge distributions[18] as well as the detection of further conformational ensembles by HDX on model proteins[19] strongly suggest that the structures observed by studies *in vacuo* correspond at least to some extent to different configurations co-existing in solution.

4.4.5 Influence of Proton Location on Thermal Stability of the Trp Cage Fold

The observations recorded here on the Trp cage miniprotein support the idea that a link between solvated protein conformations and the structures of electrosprayed ions can be established by appreciating the relative strengths of the forces acting in such an environment. Ionic polypeptides in a vacuum do not behave in a truly ergodic fashion over the studied temperature range. Furthermore, those molecules that are "loosely structured" upon desolvation proceed to collapse into much more compact gas-phase conformations. Data reported herein indicate that even $[M+2H]^{2+}$ ions are geometrically transmogrified in their solvent-free states. The small cross-sections measured for the peptide containing unprotected termini at temperatures below 500K are only matched by the most compact states sampled by simulated annealing of the zwitterionic peptide (Figure 3.6). Even though the hypothesis envisaging the presence of gas-phase $[M+(3-1)H]^{2+}$ zwitterions at low temperatures which are lost at high temperatures by proton transfer explains the observed trend, such a view requires the metastable nature of zwitterions in multiply-charged peptides to be reconciled with the extensive structural collapse that has evidently taken place. Furthermore, the existence of stable zwitterions

contradicts the mobile-proton view. The case for the perseverance of zwitterions can only be judged in the light of the K-R substituted peptides to be discussed in section 4.5.

The mobile proton view is generally consistent with the observed trends, especially with data for $[M+3H]^{3+}$ ions. The findings complicate somehow the charge assignment that can be made. As mentioned previously, Adams *et al.*[11] base a charge assignment involving a neutral K8 and a protonated Q5 for the $[M+3H]^{3+}$ ion on ECD fragment charge distributions (labelled here as WT3c). The K8-D9-R16 salt bridge is central in stabilising the Trp cage fold in an aqueous environment[2, 4, 20] implying that the K8 sidechain is charged. Three paths can be suggested for the production of such a proton localisation in the absence of more decisive data:

a) K8-Q5 proton transfer;
b) existence of a $Q5^{+}$-$K8^{+}$-$D9^{-}$-$R16^{+}$ parent ion with subsequent proton transfer from K8 to D9;
c) proton transfer from secondary charged sites to $Q5^{+}$ (yet such an event requires K8 to remain neutral throughout).

Assuming the existence of native-like structures directly after desolvation of $[M+3H]^{3+}$ ions, monitoring the distance between Q5 (C_{δ}) and K8 (N_{ζ}) (Figure 4.9) can show whether an interaction between the side-chains of these residues takes place *in vacuo*. This interaction is found to be one of the most persistent in most simulations, absent only in the case of WT3b and WTc3a, in which both side-chains are protonated. This result is hardly surprising as both residues are polar and situated on the same 'face' of the N-terminal helix and their interaction also serves to stabilise it. Given that an N-terminal helix is preserved *in vacuo* – indeed it features in the lowest-energy structures sampled by simulated annealing as shown by Figures 3.5 & 3.9 – and only one carries a charge, the side-chains of K8 and Q5 are very likely to adopt a configuration allowing proton transfer to occur between them. However, even though transfer between K8 and Q5 could happen, the opposite may also take place, and is indeed more likely considering that the gas-phase basicity of isolated lysine is higher than that of glutamine by ~7.6kcal/mol[21]. Therefore, for the Q5 side-chain to remain protonated, the N-terminal helix must be

significantly destabilised. This does in fact happen in our simulations of constructs containing free termini and a charged glutamine (WT2b, 3b and 3c), as seen in Figures 4.10 and 4.11, and understandably so, as incorporation of a positive charge in a helical stretch can destabilise the helix downstream in sequence[22]. Yet it is important to stress that in the simulations reported here the Q5-K8 interactions may also persist despite a certain destabilisation of the helix. Therefore, even though protonation at the Q5 side-chain instead of K8 does not seem entirely implausible it is not directly supported by the simulations and cannot be distinguished from experimental data.

4.4.6 Thermal Unfolding of Secondary Structural Elements

In contrast to the case in solution, secondary structure typical of the Trp cage fold, in particular helical regions, persists *in vacuo* even after loss of tertiary structure (as one may deduce from Figures 4.10 and 4.11). In terms of the temperature dependence of secondary structure of the peptide, all simulated constructs display some common trends. The N-terminal helix becomes gradually destabilised, melting of the helix close to its ends being particularly evident. However at high temperatures propagation of 3_{10} helices and even the appearance of helical turns at other locations of the peptide (e.g. in the case of trajectories for WT2z-b and WT3a) can also take place. Helices also become more dynamic at higher temperatures; furthermore a transition between α-helices and 3_{10} helices also occurs. The greater stability of 3_{10} helices compared to an α-helical conformation in isolated peptides has been demonstrated experimentally[23] and their prominence in these gas-phase simulations is encouraging; however as hydrogen-bonding interactions in the force field employed here does not involve any constraints on the angle of interacting groups one cannot expect that these simulations are capable of replicating the α-helix to 3_{10} helix transition reliably. Usage of a polarisable force field in future studies also incorporating more sophisticated potentials may improve the situation in future studies.

Another feature of high temperature trajectories is the frequent appearance of turns along the peptide sequence as well as loss of the ones present in the solution fold. Such a

change represents the loss of tertiary structure as the topology of folded Trp cage is controlled by turns that form between residues 9 and 16. The turn involving G10 manifests itself in most low-temperature trajectories, yet this is one aspect of the structure that is not captured by gas-phase simulated annealing of different constructs (Figure 3.5). In addition to a turn appearing in the originally PPII region (residues 17-19) in simulations of WT2a, WT2z-b and WT3b, an intriguing transition is that of two turns forming at intermediate temperatures (observed for WT2b, 2z, 2z-b, 3b and 3c); one involving residues 9-10 and a second one around residues 13-14. These two turns signify the re-orientation of different portions of the peptide with respect to each other giving rise to topologies reminiscent of these sampled by simulated annealing once conformational changes have occurred to the solution conformation.

Charge localisation does affect the stability of secondary structural elements. As already mentioned, a charged glutamine in both 2+ and 3+ charge states destabilises the helix downstream (i.e. at the glutamine's N-terminal end, the charge being located near the C-terminal end of the helical region). A similar destabilisation is also observed when the N-terminus is protonated, in the case of WT2z-b, 3a and 3c. Due to the intrinsic ability of helices to align backbone NH and CO groups, a long-range macro-dipole is formed; incorporation of a proton at the N-terminal end destabilises this macro-dipole and perturbs helical stability. Conversely, a positive charge situated at the C-terminal end of a peptide segment will stabilise the macro-dipole and, consequently, helical conformations[22]. Furthermore, a proton located on a flexible side-chain of a C-terminal residue, such as a lysine residue, hydrogen-bonding between the charged group and the exposed backbone carbonyl groups (which at the end of a helix do not form any hydrogen bonds) significantly stabilises the helix.

Protonation of the N-terminus may stabilise the overall topology of the solution fold, yet it concurrently effects a partial 'melting' of the helix. However, due to the presence of basic residues further downstream as well as inter-residue hydrogen-bonding between helix constituents, such as the Q5-K8 interaction depicted in Figure 4.9, the N-terminal helix survives, at least in part, in these solvent-free Trp cage ions.

Addition of terminal groups stabilises the N-terminal helix (Figures 4.10 & 4.11); the N-terminal acetyl cap can form an additional hydrogen bond with backbone NH of Y3. α-helical character is most stable in MD trajectories of peptides with protected termini. Nevertheless, the existence of capping groups does not hinder WTc2 of forming an interaction between termini; interestingly, WTc2 is also the capped construct which preserves traits of the solution fold most faithfully (Figure 4.6). WTc3a and WTc3b do preserve their helical character to a greater extent than all uncapped 3+ systems simulated (Figure 4.11); yet this feature does not hinder loss of solution-like overall backbone topology. Therefore we can conclude that in the absence of water co-operativity of peptide folding is lost, as solvophobic effects are no longer significant; yet, providing a sufficiently "gentle" process is employed for desolvating the peptides, solution conformations can be preserved, if the activation barrier for the conformational change in the gas-phase PES is higher than the energies available to the ion in the ion mobility mass spectrometer. Temperature-dependent variation of calculated cross-sections for capped constructs compares favourably with experimental trends but deviates significantly at a quantitative level. Protection of the termini reduces charge location possibilities rendering these simulations more representative. Therefore, strong interactions involving peptide termini are directly implicated in producing the compact structures observed experimentally for free-termini peptides. In contrast, peptides with protected termini undergo a structural collapse to a lesser extent (Figure 4.3) and are very likely to preserve secondary structural elements and the overall topology of the Trp cage fold in the gas phase.

As charge delocalisation appears to be a significant factor in determining isolated peptide structure, studies on K-to-R and *vice versa* substituents may illuminate the structural characteristics of the miniprotein in the absence of water.

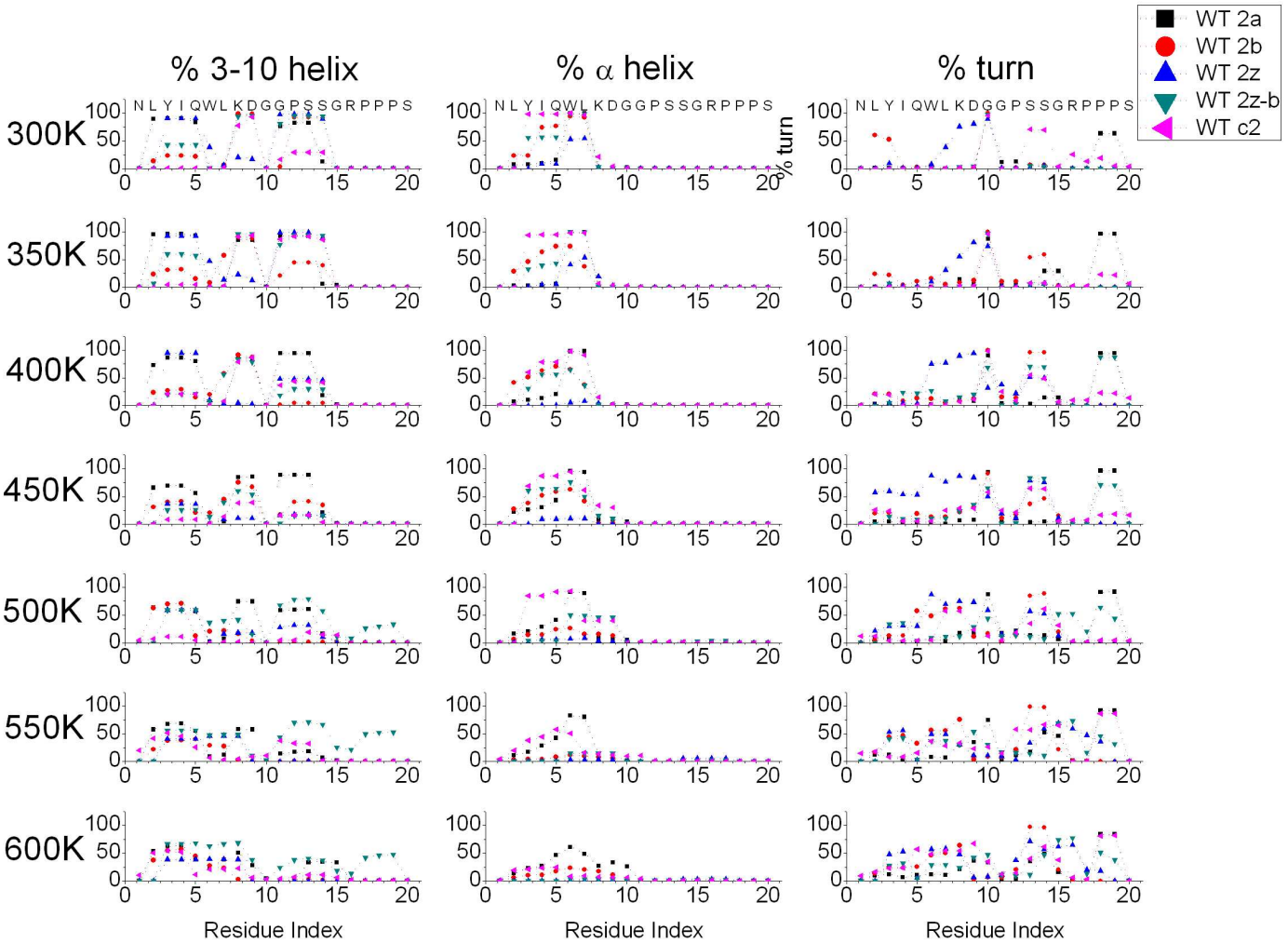


Figure 4.10
Average helical content and turn character, calculated with the DSSP model, of $[M+2H]^{2+}$ 'charge isomers' of WT Trp cage containing either unprotected or protected termini resulting from MD simulations at various temperatures. Averages are displayed as a function of peptide sequence. Left column, 3_{10} helical character; central column, α-helical character; right column, turn character.

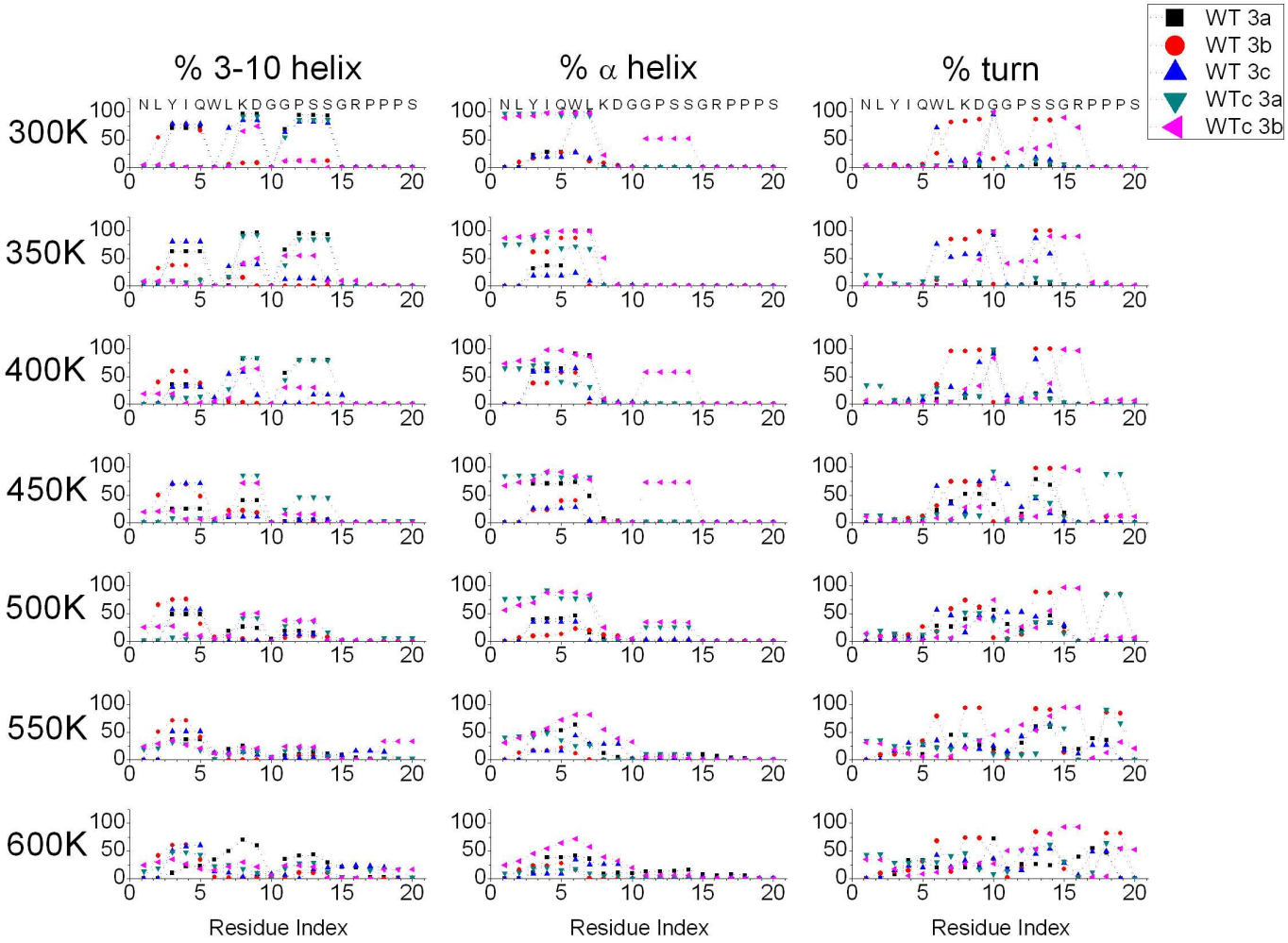


Figure 4.11
Average helical content and turn character, calculated with the DSSP model, of $[M+3H]^{3+}$ 'charge isomers' of WT Trp cage containing either unprotected or protected termini resulting from MD simulations at various temperatures. Averages are displayed as a function of peptide sequence. Left column, 3_{10} helical character; central column, α-helical character; right column, turn character.

4.5 Effect of Arg-Lys substitutions on Trp cage's Gas-phase Conformation

4.5.1 Thermal Unfolding of Ionised Trp Cage Arg & Lys Variants as Probed by IMMS

The addition of protecting groups at the peptide termini of the Trp cage miniprotein has a much more readily detectable (by IMMS) effect on the structure of the solvent-free ions than Arg-Lys substitution. Experimental results are summarised in Figure 4.12. Similarly to WT, $[M+2H]^{2+}$ ions (Figure 4.12a) of the remaining three constructs also exhibit a very modest increase in collision cross section between 300 and 550K[*]. A slight unfolding can be observed for K8R between 350 and 400K (a difference in cross section of 35Å^2 is observed, yet it must be noted that a large uncertainty is associated with these measurements); a similar transition may occur for R16K over the 350-450K range. DM cross sections appear to fluctuate around 385Å^2, the only exception being the measurement taken at 400K. Unfortunately these fluctuations are dwarfed by estimated error, obscuring any possible interpretation. Likewise, $[M+3H]^{3+}$ ions of free-termini peptides follow a similar course (Figure 4.12b): gradual increase with temperature, reaching geometries slightly more extended than the solution fold at the highest temperatures. WT appears to 'soar' above other constructs up to 450K. Interestingly, R16K is more extended at 450 and 500K – yet the reliability of this difference is unclear. A much greater difference is observed between WT and R16K – both containing protected termini. As discussed above, incorporating capping groups results in more extended gas-phase structures for both $[M+2H]^{2+}$ and $[M+3H]^{3+}$ ions, also inducing an unfolding transition to occur in $[M+3H]^{3+}$ ions at low temperatures. Substitution of Arg16 with Lys reduces the overall structural stability of both charge states, the effect being more severe for $[M+2H]^{2+}$ ions (Figure 4.12c). Proton sequestration by R16 reduces the occupation of secondary charged sites, thereby improving the stability of the peptide. We may therefore conclude that such alternative proton locations have a

[*] N.B. Data at 550K correspond to single measurements and were collected using commercially available nano-ESI tips (PROTANA) as opposed to capillaries pulled in-house used for the remainder of the IMMS experiments: for these reasons they must be deemed less representative.

destabilising effect on the structure, a conclusion supported by simulation of capped $[M+3H]^{3+}$ ions (see sections 3.8.5, 4.4 and discussion bellow). The flexible nature of lysine's side-chain, especially of the amine group compared to the guanidino group of arginine, the former potentially allowing a greater number of possibilities for hydrogen-bonding interactions, may also account for the observed effect.

If at all different, WT, K8R and R16K ions are expected to differ in their relative populations of different charge locations. Models assuming 'static' protons must not reveal any significant difference between them: and this is exactly what *in silico* studies imply. MD simulations, starting from the NMR configuration, were performed for all constructs with unprotected termini as well as an additional D9N variant, which was not studied experimentally but has been investigated by other authors using different methodologies[6, 8]. For practical reasons only three different charge assignments were considered; 2a, 2z and 3a (as D9N lacks an aspartic acid, only 2a and 3a were simulated for that construct). Trajectories generated for K8R, R16K and DM produced similar results to equivalent charge assignments for WT, as far as the general trends pertaining to thermally-induced conformational fluctuations are concerned. Parallel behaviour between constructs as studied by simulated annealing has also been observed (section 3.8.5), supporting the hypothesis that moderate variations in the structure of the basic residue exert a small influence on peptide geometries given that proton location in peptide sequence is the same. Consequently it is probable that little extra information is lost by omission of specific charge variants from our simulation studies.

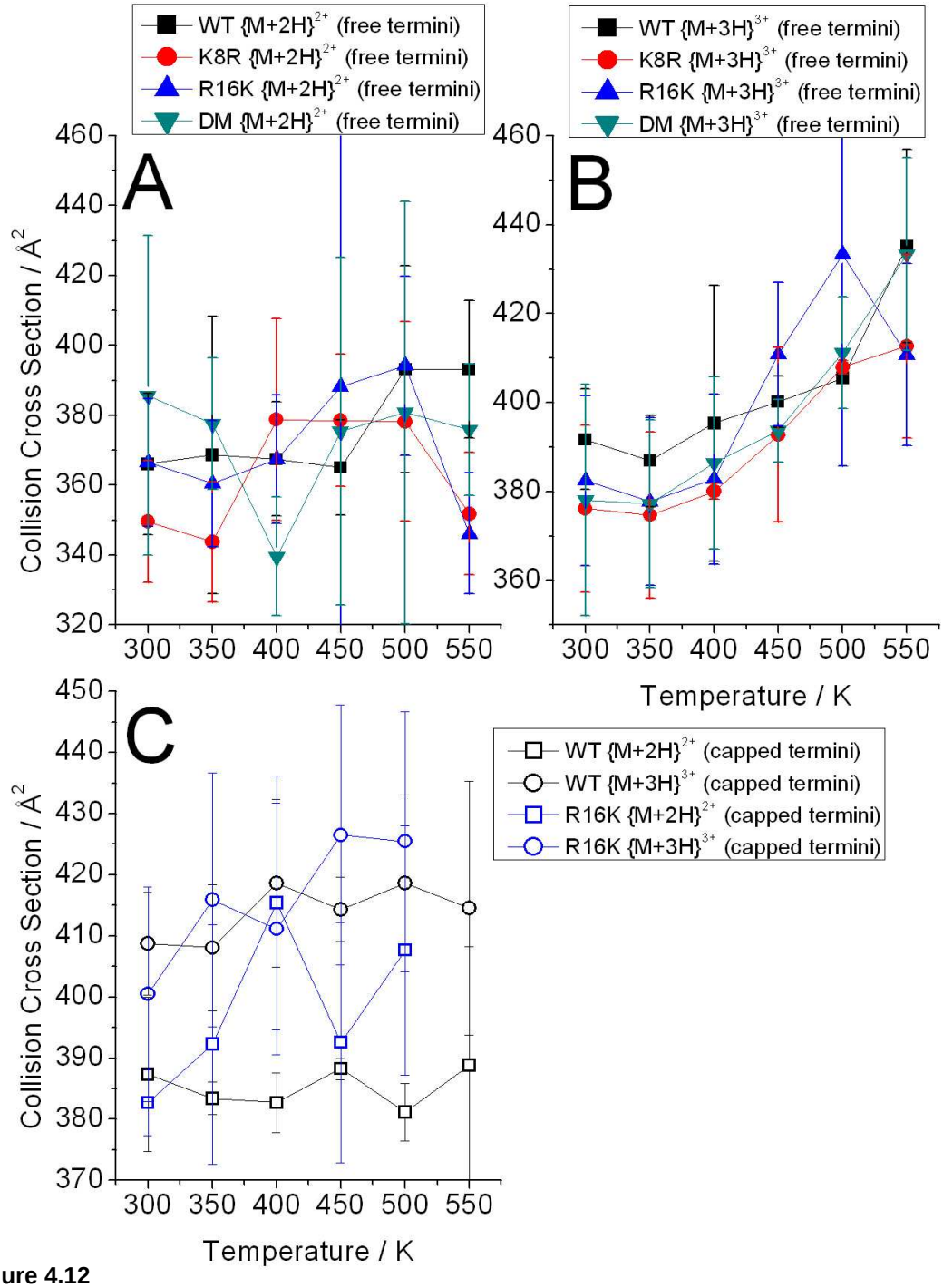


Figure 4.12
Collision cross sections of Trp cage constructs at elevated temperatures determined by IMMS. a) $[M+2H]^{2+}$ ions and b) $[M+3H]^{3+}$ ions of peptides containing unprotected (free) termini. Black squares, WT; red circles, K8R; blue triangles, R16K; green inverse triangles, DM. c) collision cross-sections of peptides with protected termini. Squares: $[M+2H]^{2+}$ charge states; circles: $[M+3H]^{3+}$ charge states; black symbols: WT; blue symbols: R16K (cross-section ranges are *not* shared among all graphs).

4.5.2 MD Simulations of Gas-phase Ionised Arg-Lys Variants of Trp Cage at High Temperatures

The Trp cage fold remains recognisable in all trajectories up to a simulation temperature of 400K, with the exception of D9N which undergoes substantial changes at the onset of the simulation at 300K (Figure 4.15). Thus even the presence of a 'neutral' carboxyl group enhances the stability of the solution fold. In agreement with the trends resulting from simulations of WT charge isomers discussed in the previous section, the presence of an anionic carboxylic acid or a persistent interaction between the termini both stabilise the Trp cage fold. A particularly interesting occurrence of the latter phenomenon can be seen in the case of R16K 2a. In contrast to other simulations presented in Figure 4.15, this trajectory undergoes a smaller conformational rearrangement at 350K and another one at 550K, the backbone RMSD to the NMR structure remaining below 3Å until the latter event; as Figure 4.16 clearly displays, the termini interact for most of this interval. Conversely the solution fold is less persistent during WT 2a MD; however the cause of this event as well as its significance cannot be decided from these studies alone. As noted above, conformational changes proceed via extended intermediates but often result to more compact structures (Figures 4.13 and 4.14), excepting trajectories at a temperature ≥550K, supporting the idea of a structural collapse occurring in gas-phase protein ions not bearing a large number of charges.

The helix-stabilising Q5-X16 interaction (displayed in Figure 4.17) is more persistent in constructs where X=R (i.e. K8R and DM) due to the greater number of polar hydrogen atoms, more extended and relatively rigid geometry of arginine's guanidino group compared to lysine's amine. In terms of secondary structure content, constructs of similar charge assignment differ mainly in their relative content of 3_{10} and α-helical elements as judged by the DSSP model (Figures 4.19, 4.20, 4.21). Helix 'melting' is observed at higher temperatures, with the concurrent emergence of miscellaneous turns. A turn around G10 is observed in all trajectories in which the topology of the Trp cage fold is preserved. Yet a small number of 'outliers' are also evident. Between 350-450K DM 2a forms two turns: one including residues 8-10 and a second incorporating residues

13-14. WT 2a and D9N 2a form a turn around prolines 17 & 18. Similar turns are recognised in other constructs at higher temperatures, indicating that these are diagnostic of the collapse, or at least the onset of a collapse, from the folded state of the solvated molecule into a gas-phase structure. The similarity of the secondary structure content of D9N with that of other simulated Trp cage variants can be taken as additional evidence of the stability of hydrogen-bonded structures in the gas phase, trans-conformational events in small proteins in such an environment being driven by the loss of hydrophobic interactions.

4.5.3 Comparison between IMMS and MD Studies of Arg-Lys Substituents

Overall, MD simulations only reveal small and generally insignificant differences between variants containing a similar charge assignment. Additionally, only minor differences between constructs with unprotected termini are detected by IMMS at high temperatures. The far greater effect of terminal capping groups on measured collision cross-sections implicate the N-terminus as a charge-bearing moiety in $[M+3H]^{3+}$ ions when such groups are absent. Yet the temperature resistance of $[M+2H]^{2+}$ ions in contrast to $[M+3H]^{3+}$ species, the all-too-evident structural collapse that has taken place, as well as the stabilising effect of a strong electrostatic interaction between the termini imply that intramolecular proton transfer does indeed take place, at least in $[M+3H]^{3+}$ ions, an event which is likely to mediate gas-phase conformational changes at the millisecond timescale. Such proton 'scrambling' would also suggest that two protons reside in the most basic groups, that is arginine and lysine side-chains.

The similarity in the temperature dependence of WT and R16K experimental cross-sections compared to K8R and DM cross-sections imply that proton localisation, at least at elevated temperatures, is also similar, rendering Q5 an unlikely side-chain to sequester a proton.

The small cross-sections obtained for K8R $[M+2H]^{2+}$ at 300 and 350K, as well as the transition to more extended geometries occurring between 350 and 400K may support the presence of a salt bridge in this ion population. However, measured cross sections are much smaller than the NMR geometry, leading to the conclusion that major

rearrangements have taken place, which, given the metastable nature of salt bridges in multiply-charged unsolvated cations, contradicts the salt-bridge view. Moreover, there is no evidence in the experimental ATDs of two ionic populations corresponding to the salt-bridge and charge-solvated species, indicating that if neutralisation of the salt bridge takes place then it may occur at time-scales similar to that of the experiment. Incorporation of cold gas in the source region and IMMS experiments at temperatures lower than room temperature (colder buffer gas will also increase the resolution of the drift tube) may aid to resolve the salt-bridge dilemma. Gas-phase spectroscopic techniques could also be enlisted in resolving this problem for the Trp cage miniprotein, as it has done before for smaller gas-phase peptides[24].

This combined experimental and computational effort on the Arg-Lys substitution variants of Trp cage composes a picture of gas-phase protein structure that contrasts the one in solution. CD experiments (discussed in section 3.8.1) suggest that a choice of more basic residues for X8 and X16 strengthens the salt-bridge including D9, thus stabilising the fold. Furthermore, phenomena which rely on rapid proton transfer, such as gas-phase fragmentation by collisional activation (section 3.8.2) and rapid exchange rate of gas-phase HDX (Table 3.6) also correlate with gas-phase basicity of the individual constituent amino acids for peptides with capped termini. Gas-phase IMMS does indeed indicate greater 'unfolding' for capped R16K $[M+2H]^{2+}$ compared to analogous state of WT, which is in general agreement with these findings, implying that the correlation between proton location and structure is not merely an artefact arising from over-estimating partial charges in molecular mechanics simulations. Yet, in the absence of terminal protecting groups these differences are greatly obscured. Consequently, interactions involving the termini take up a major role in determining gas-phase structural stability of peptides. On the basis of these observations it could be predicted that the trends observed by CID and HDX for capped $[M+2H]^{2+}$ ions will be less clear for free-termini $[M+2H]^{2+}$ ions. Gas-phase free-termini peptide cations adopt very compact structures exhibiting moderate extension even at high temperatures. Simulations and experiments indicate that electrostatics dominate gas-phase geometries, with the weaker forces, significant in an aqueous environment, becoming much less

influential. Study of the d-Pro12 variant, which does not fold efficiently in solution, provides further support for this point.

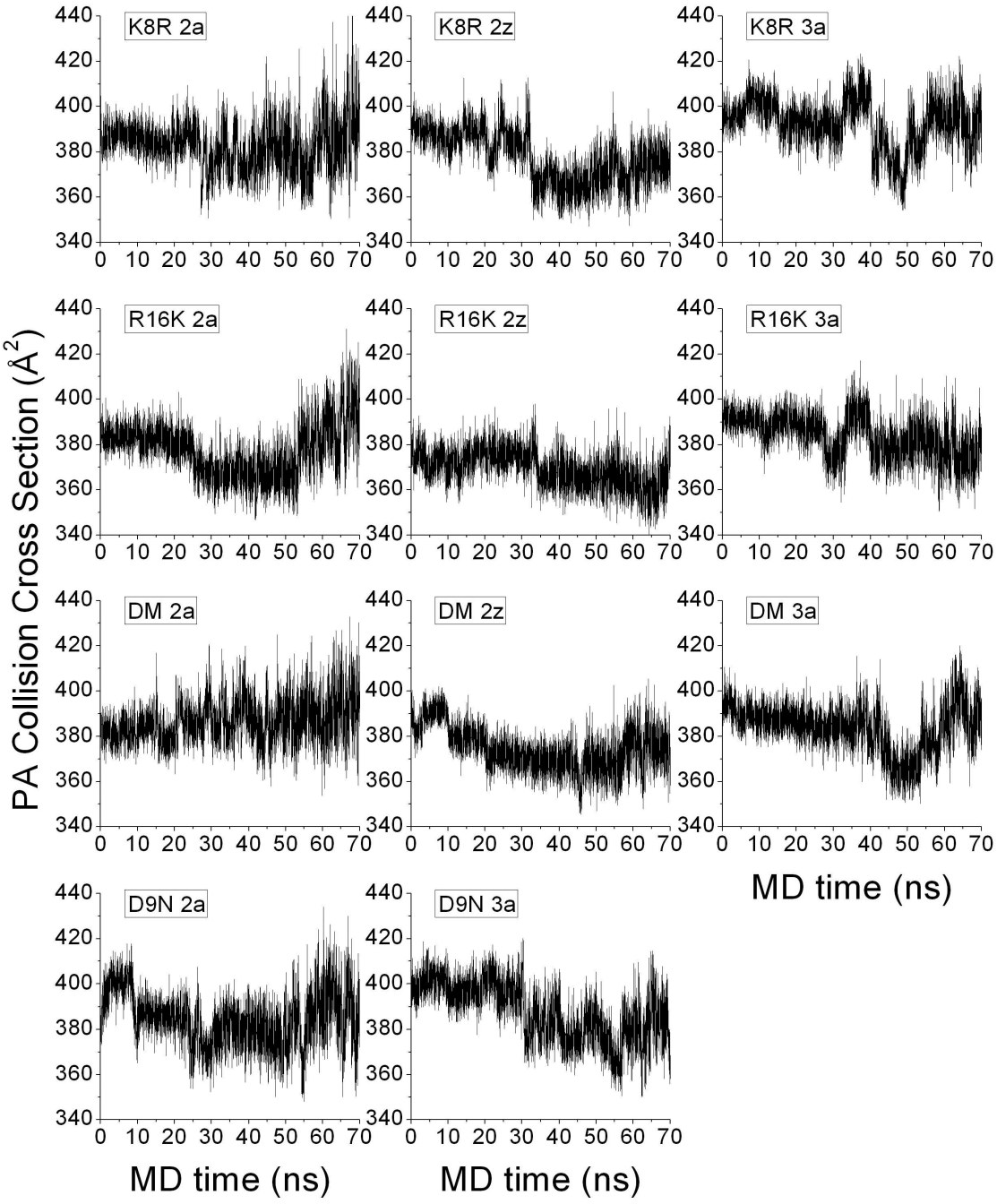


Figure 4.13
Collision cross sections for K8R, R16K, DM and D9N variants 2a, 2z and 3a, calculated using the projection approximation method. Cross sections were estimated every 10ps of simulation time. Simulation temperature increments by 50K every 10ns of the trajectory: 0-10ns: 300K; 10-20ns: 350K; 20-30ns: 400K; 30-40ns: 450K; 40-50ns: 500K; 50-60ns: 550K; 60-70ns: 600K. All graphs are in scale.

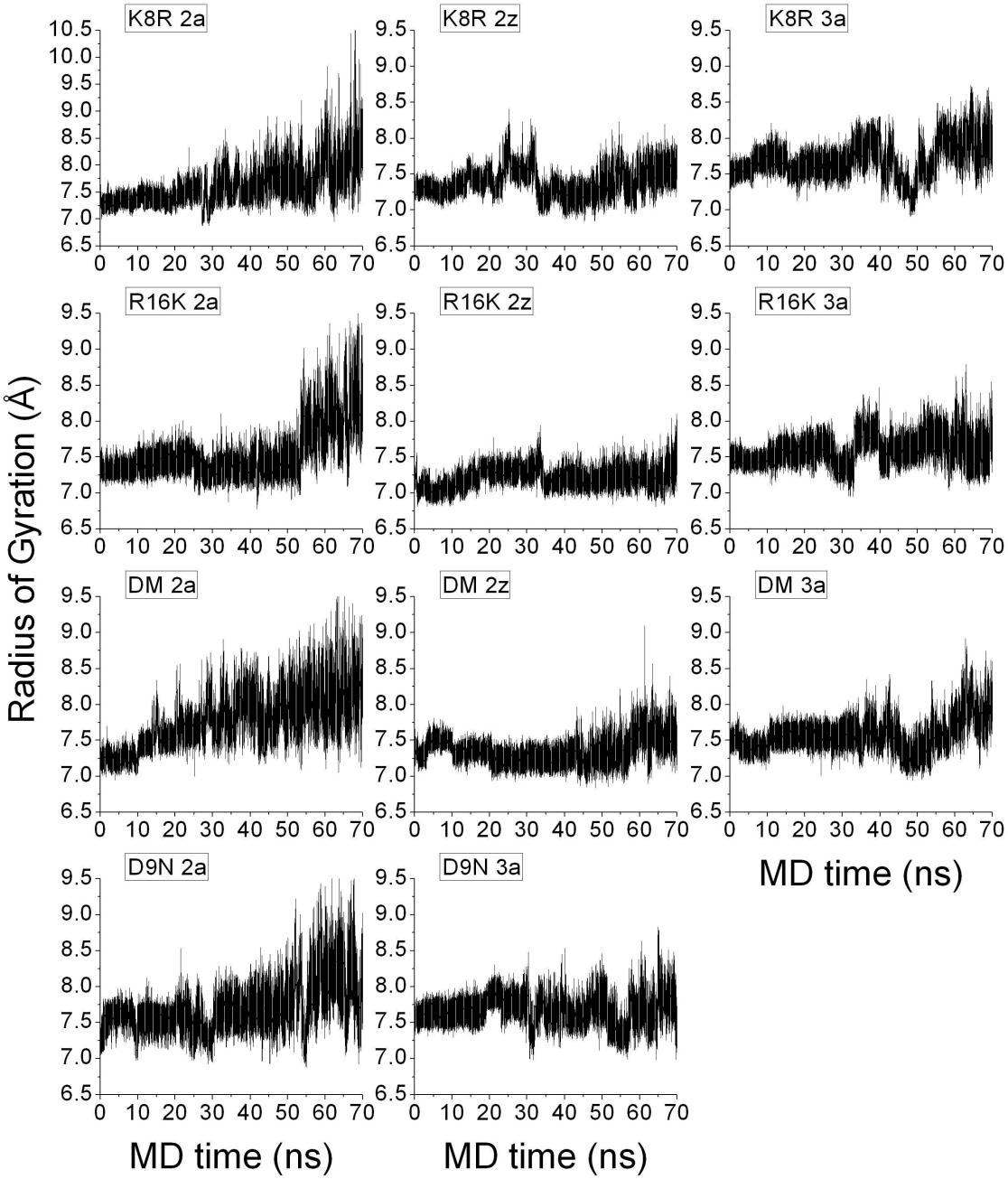


Figure 4.14
Radii of gyration for MD simulations of K8R, R16K, DM and D9N variants 2a, 2z and 3a. Simulation temperature increments by 50K every 10ns of the trajectory: 0-10ns: 300K; 10-20ns: 350K; 20-30ns: 400K; 30-40ns: 450K; 40-50ns: 500K; 50-60ns: 550K; 60-70ns: 600K. All graphs are *not* in scale.

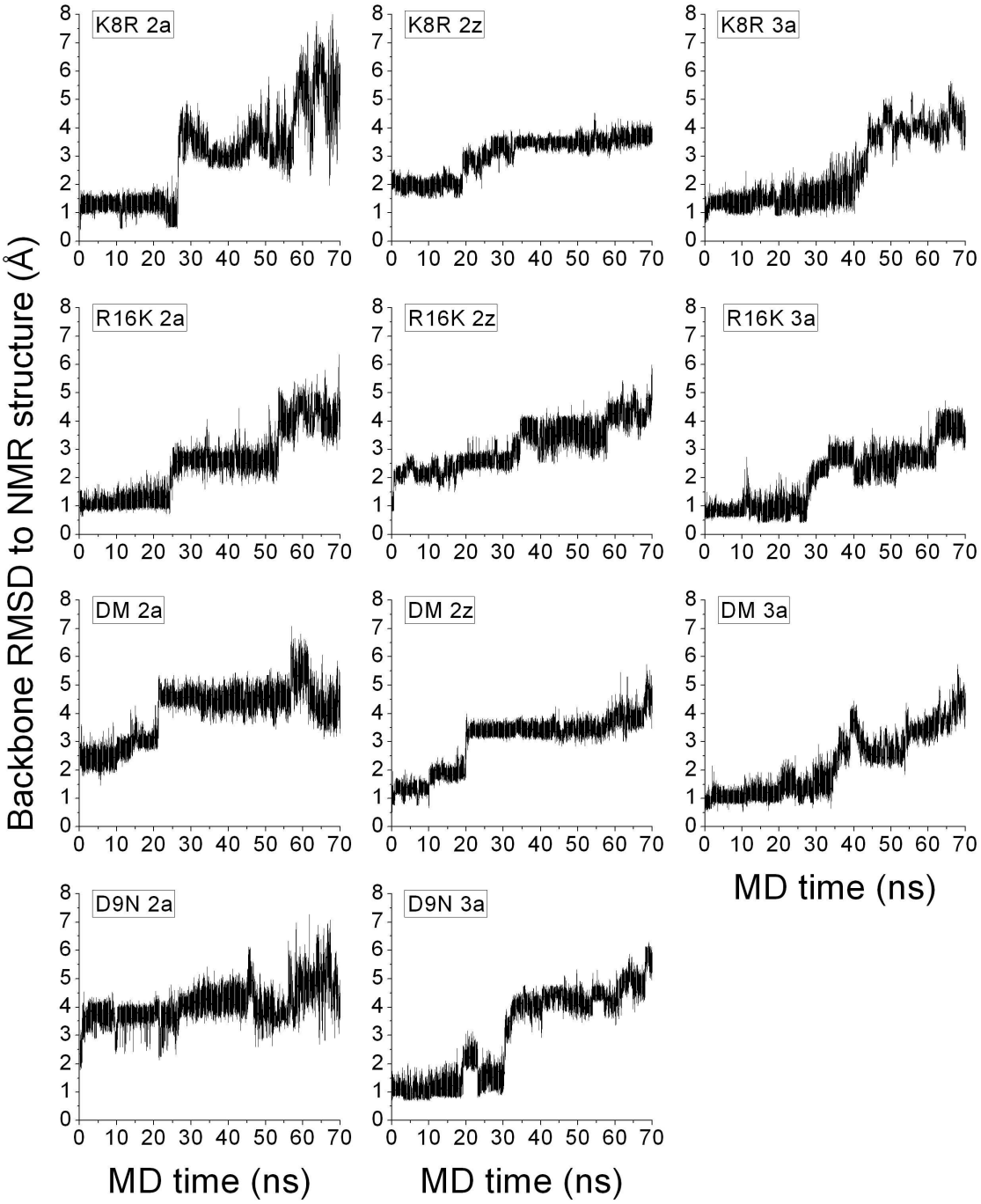


Figure 4.15
Heavy-atom backbone RMSD calculated for structures generated by MD simulations of K8R, R16K, DM and D9N variants 2a, 2z and 3a. Simulation temperature increments by 50K every 10ns of the trajectory: 0-10ns: 300K; 10-20ns: 350K; 20-30ns: 400K; 30-40ns: 450K; 40-50ns: 500K; 50-60ns: 550K; 60-70ns: 600K. All graphs are in scale.

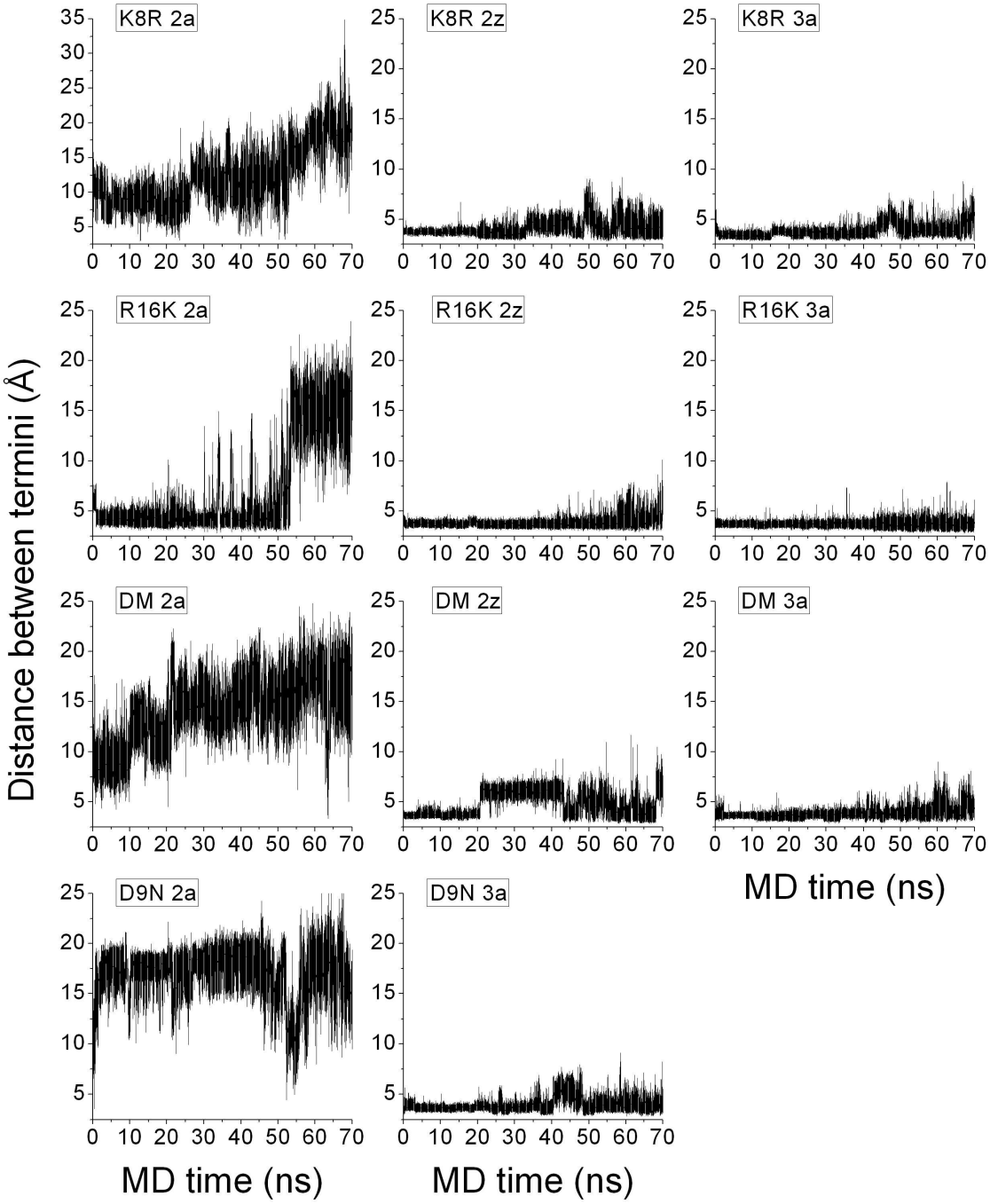


Figure 4.16
Distance between backbone nitrogen of Asn1 and backbone carbonyl carbon of Ser20, calculated from MD trajectories of K8R, R16K, DM and D9N variants 2a, 2z and 3a. Simulation temperature increments by 50K every 10ns of the trajectory: 0-10ns: 300K; 10-20ns: 350K; 20-30ns: 400K; 30-40ns: 450K; 40-50ns: 500K; 50-60ns: 550K; 60-70ns: 600K. All graphs are *not* in scale.

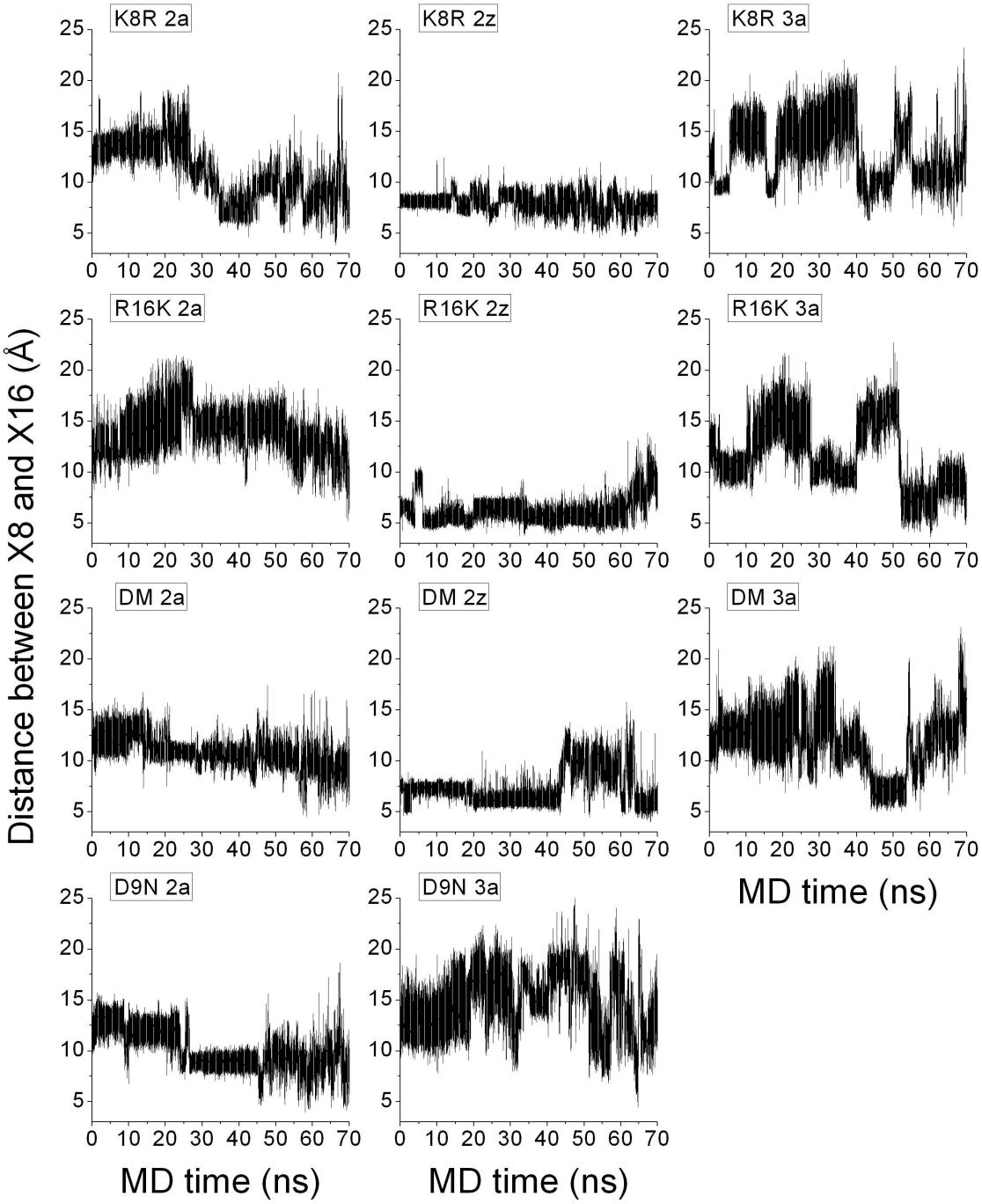


Figure 4.17
Distance between the side-chains of the basic residues at position 8 and 16 for MD simulations of K8R, R16K, DM and D9N variants 2a, 2z and 3a. In the case of lysine, the side-chain nitrogen (N_ζ) was chosen as a reference atom whereas for arginine the carbon at the centre of the guanidino group (C_ζ) was selected as reference. Simulation temperature increments by 50K every 10ns of the trajectory: 0-10ns: 300K; 10-20ns: 350K; 20-30ns: 400K; 30-40ns: 450K; 40-50ns: 500K; 50-60ns: 550K; 60-70ns: 600K. All graphs are in scale.

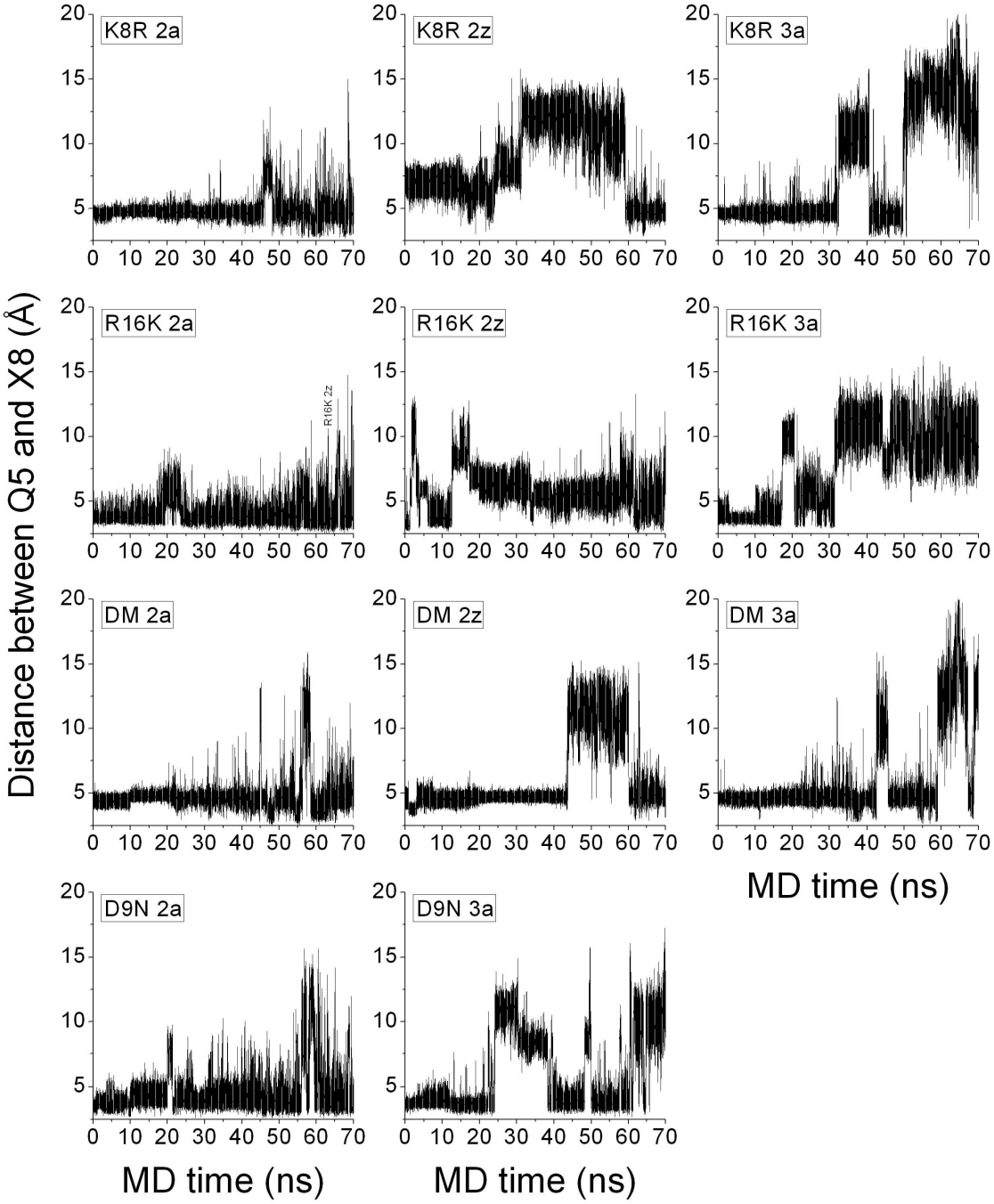


Figure 4.18
Distance between C_δ of Gln5 and the side-chain of the basic residue at position 8 for MD simulations of K8R, R16K, DM and D9N variants 2a, 2z and 3a. For lysine N_ζ serves as the reference atom whereas for arginine this role befalls on C_ζ. Simulation temperature increments by 50K every 10ns of the trajectory: 0-10ns: 300K; 10-20ns: 350K; 20-30ns: 400K; 30-40ns: 450K; 40-50ns: 500K; 50-60ns: 550K; 60-70ns: 600K. All graphs are in scale.

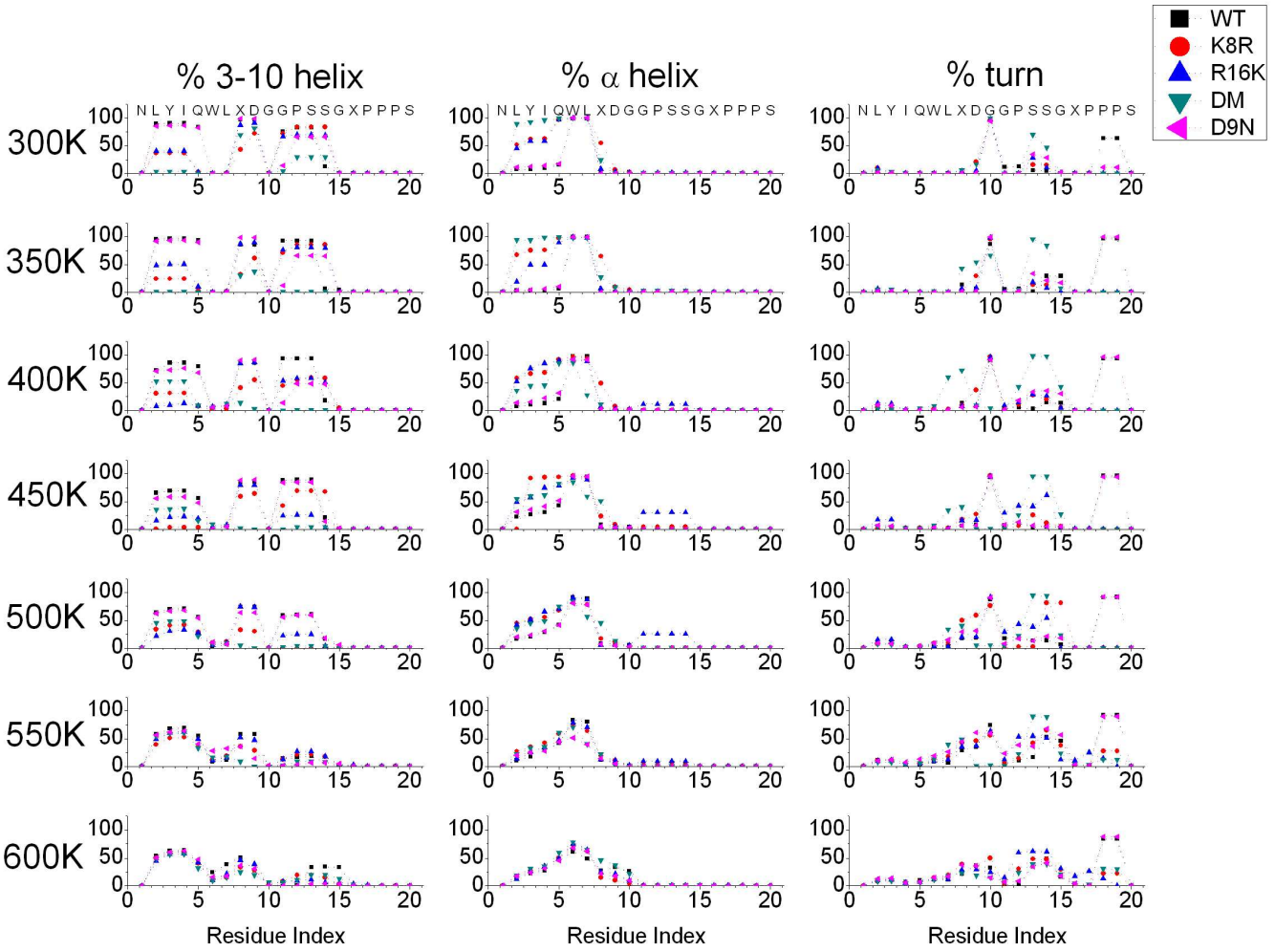


Figure 4.19
Average helical content and turn character for structures generated by MD at various temperatures of charge state 2a of WT (black squares), K8R (red circles), R16K (blue triangles), DM (green triangles) and D9N (pink triangles), calculated using the DSSP model. All constructs contain free termini. Averages are displayed as a function of peptide sequence. Left column, 3_{10} helical character; central column, α-helical character; right column, turn character.

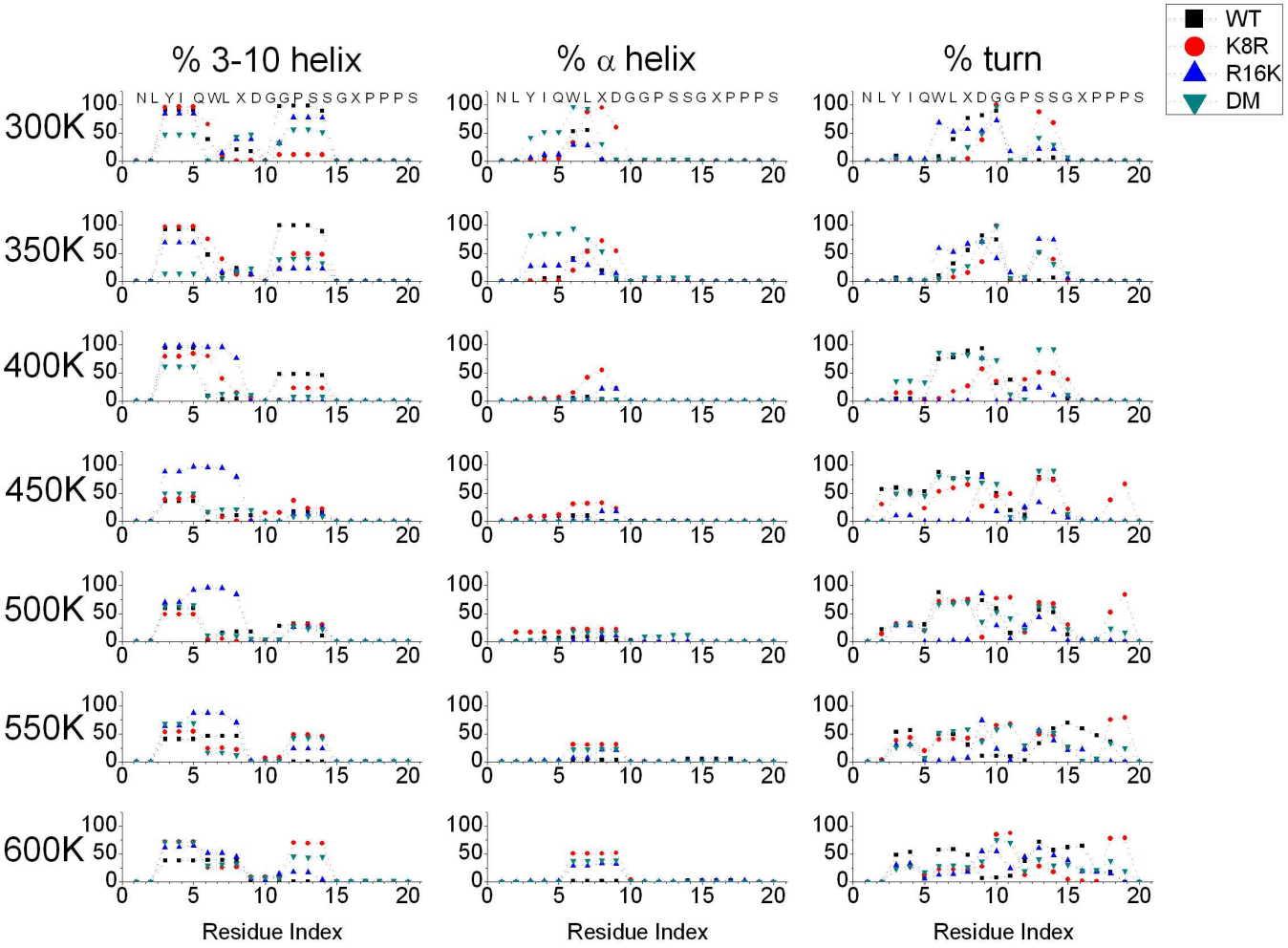


Figure 4.20
Average helical content and turn character for structures generated by MD at various temperatures of charge state 2z of WT (black squares), K8R (red circles), R16K (blue triangles), DM (green triangles) and D9N (pink triangles), calculated using the DSSP model. All constructs contain free termini. Averages are displayed as a function of peptide sequence. Left column, 3_{10} helical character; central column, α-helical character; right column, turn character.

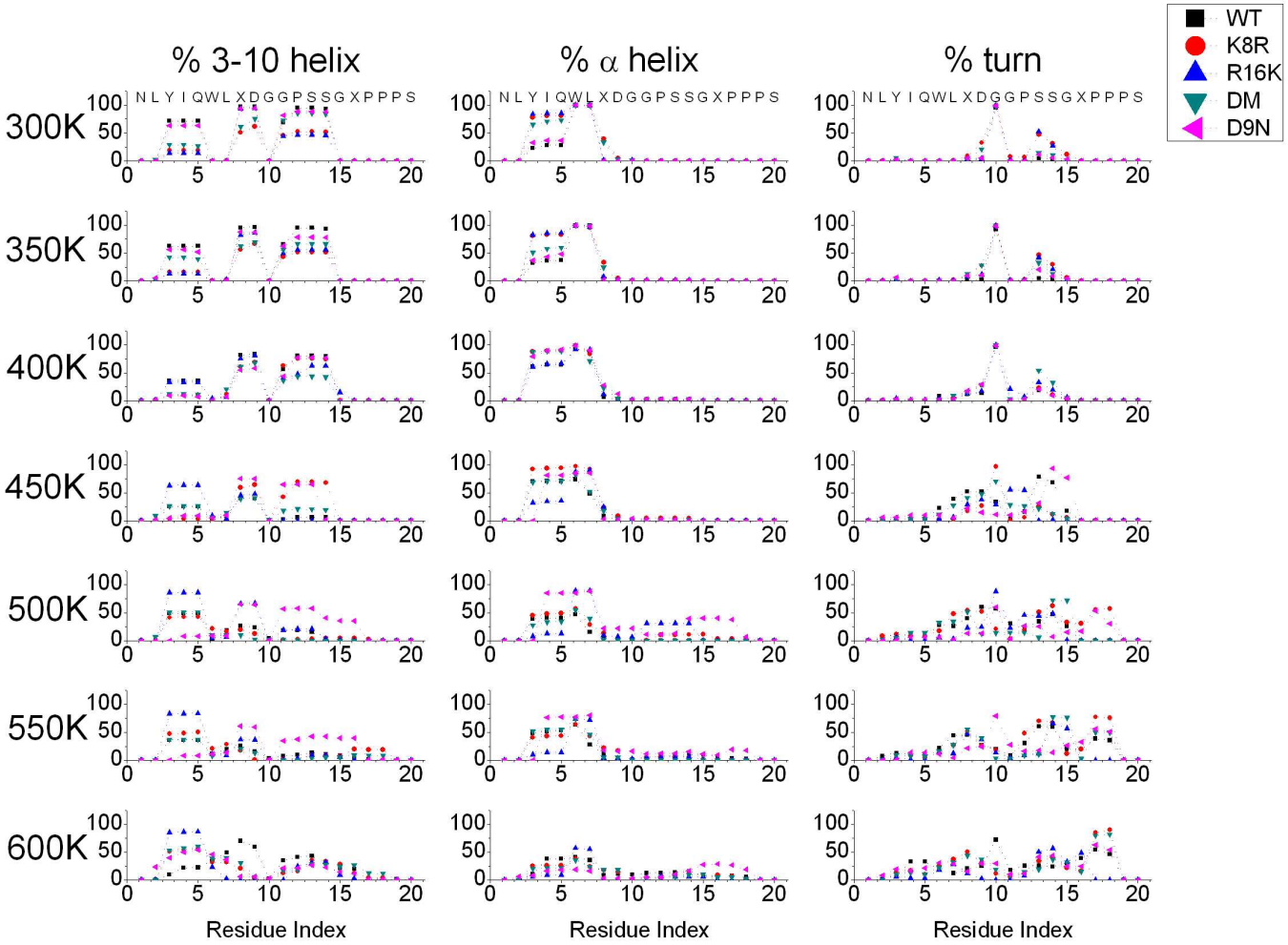


Figure 4.21
Average helical content and turn character for structures generated by MD at various temperatures of charge state 3a of WT (black squares), K8R (red circles), R16K (blue triangles), DM (green triangles) and D9N (pink triangles), calculated using the DSSP model. All constructs contain free termini. Averages are displayed as a function of peptide sequence. Left column, 3_{10} helical character; central column, α-helical character; right column, turn character.

4.6 Effect of a d-Proline at Position 12

4.6.1 Effect of d-Pro12 Attested by IMMS

Studies on the heterochiral d-Pro12 Trp cage construct set up a remarkable contrast between conformational characteristics of solvated and unsolvated ions. As described in section 3.9, d-Pro12 does not appear well-folded in solution; a comparison of ambient-temperature CD spectra of d-Pro12 with high temperature spectra obtained for WT (Figure 4.2) at T=85°C (358.15K) supports this point. Essentially, over both the far- and near-UV range, d-Pro12 at ambient temperature gives rise to dichroic signatures very similar to those obtained for thermally-unfolded WT. Nevertheless, the differences between the constructs *in vacuo* are less pronounced.

As one can see in Figure 4.22, trends of cross-section against temperature for the two constructs are similar for both $[M+2H]^{2+}$ and $[M+3H]^{3+}$ ions. Collision cross-sections for d-Pro12 $[M+2H]^{2+}$ ions show a more sustained increase with temperature whereas equivalent WT species only become more extended at 500 and 550K; in the case of d-Pro12 the change at these temperatures is less dramatic. d-Pro12 $[M+3H]^{3+}$ ions vary with temperature in a fashion nearly indistinguishable to WT save the fact that the size of the former is consistently lower for all temperatures other than 300K where the opposite is true. For both $[M+2H]^{2+}$ and $[M+3H]^{3+}$ states, d-Pro12 exhibits a more noticeable conformational collapse as the temperature is increased from 300 to 350K. It is important to note that all these traits lie below estimated error and ought to be accepted cautiously. The similarity between WT and d-Pro12 leads to the conclusion that conformational changes occur in desolvated, 'heteropure', Trp cage ions even at low temperatures, either due to source heating or residence of the ions in the instrument (once past the initial source elements, ions within the MoQToF spend most time in the accumulation region and the drift tube). The – admittedly very slight and not altogether reliable – differences between the gas-phase unravelling of WT and d-Pro12 Trp cage may hint to some actual differences in three-dimensional structure. In particular, conformational changes in $[M+2H]^{2+}$ WT ions occur over a narrower temperature range than d-Pro12, possibly suggesting a somewhat greater structural stability for the former.

Intriguingly, gas-phase FRET studies performed by Iavarone *et al.* [6, 8] show that D9N behaves very similarly with temperature to WT Trp cage *in vacuo* even though their differences in solution are great. On a more general level (as Iavarone *et al.* investigated a different set of Trp cage variants) the stability of the global fold in solution does not affect gas-phase unfolding; such an event can be attributed to loss of the co-operativity of the fold due to the absence of 'solvophobic' effects. This statement is corroborated by the gas-phase MD simulations.

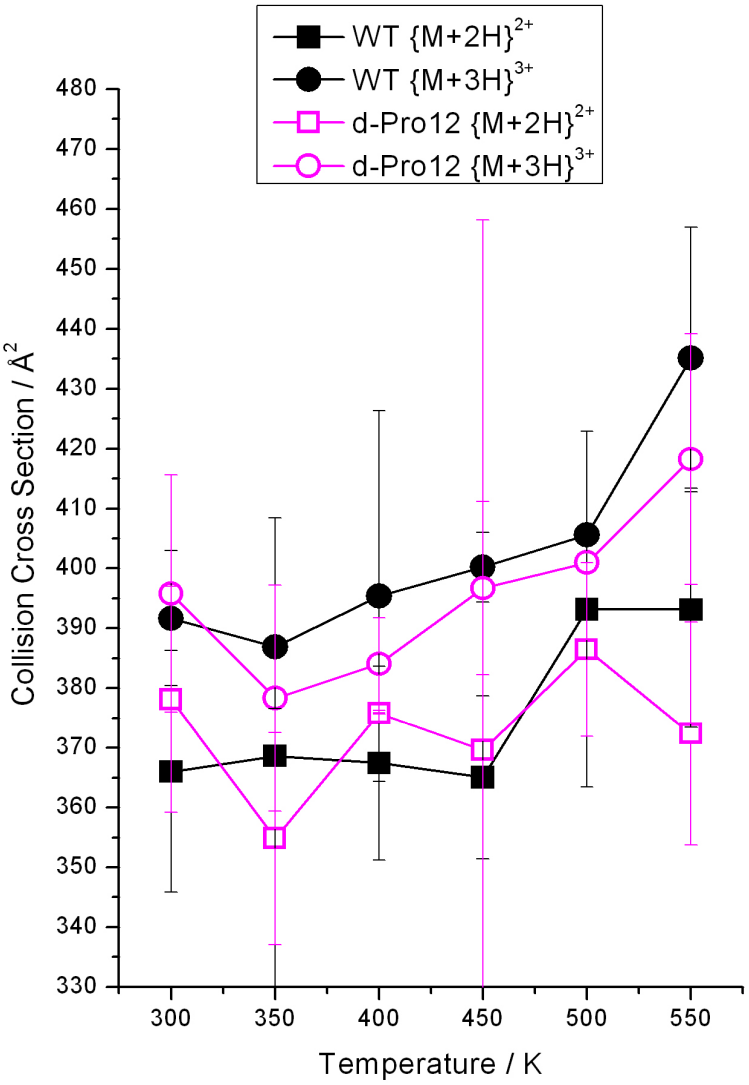


Figure 4.22
Experimentally-determined collision cross sections of d-Pro12 $[M+2H]^{2+}$ (open pink squares) and $[M+3H]^{3+}$ ions (open pink circles) as a function of temperature. Data for WT ions (black symbols; squares: $[M+2H]^{2+}$ ions; circles: $[M+3H]^{3+}$ ions) are also shown for comparison.

4.6.2 MD Studies of d-Pro12 in a Vacuum

Even though d-Pro12 adopting a 3-D topology akin to the one found for folded WT is an unlikely situation, such an endeavour can serve as a 'control' simulation, i.e. it will aid to indicate the extent of destabilisation brought about by insertion of a d-Pro in terms of the simple simulation scheme adopted here.

As for other peptide constructs, MD simulations of the d-Pro12 variant took the NMR conformation as a starting point; due to the chirality change of P12 in this system the proline side-chain 'flips out' from the core of the molecule. The effect this modification imparts on this conformation in a vacuum is noticeable, yet not extreme. d-Pro12 2a already has lost much of its similarity to the starting configuration by the time the trajectory at 300K is initiated. As with other constructs, such a gas-phase conformation change is in fact a structural collapse instead of a move to more extended configurations. The trajectory of d-Pro12 containing the K8-D9-R16 salt bridge is characterised by a mild albeit significant stabilisation of the solution fold which is recognisable up to 350K, as it can be inferred from Figure 4.23. Both simulated 2+ charge states adopted configurations more extended than the ones visited by equivalent WT constructs, a fact which may be deduced from calculated radii of gyration. Surprisingly, protonation of the N-terminus (in d-Pro12 3a) stabilises the solution configuration to a great extent over the entire simulated temperature range. The same effect has been observed for all (free-termini) peptides protonated at the N-terminus, K8 and R16; this datum shows that the interaction between the termini plays a fundamental role in preserving the Trp cage fold *in vacuo*, at least *in silico*. Disruption of the hydrophobic core achieved by altering the chirality of P12 is compensated in these simulations by introducing strong electrostatics, the ones strengthening the interaction between the termini being the most effective of the simulated set.

Analysis of secondary structures using the DSSP model[25] (Figures 4.24-4.26), in agreement with similar analyses described in preceding sections, indicate persistence of simulated 'native-like' secondary structure accompanied by gradual 'melting' with temperature. Intriguingly, helical stretches in d-Pro12 constructs survive well even at high temperatures, compared to equivalent data for WT. Helices are somewhat less

stable towards the middle region of the peptide, an effect symptomatic of conformational reorganisations in the vicinity of the d-amino acid. A notable feature of helical stability is that of residues 5 & 6. In WT these form the core of the helix and assume a stable α-helical character; yet in d-Pro12 simulations they are more flexible conformationally, especially at low temperatures.

Another important feature is this of the crucial turn forming around G10 in the solution fold. In d-Pro12 2a this configuration is only witnessed at 300K MD trajectories; at higher temperatures the turn is replaced by two more flexible turns on either side of that region. This effect is similar to what has been observed by dihedral angle analysis of d-Pro12 structures generated by simulated annealing (section 3.9.4). Similar effects are observed in the other two charge assignments, as soon as the 'native' topology collapses into a more stable gas-phase structure. Notably, protonation of the N-terminus forces the turn around G10 to remain intact even at high temperatures. Patterns of turns existing in this very construct, as well as their temperature-dependent changes, correlate between WT and d-Pro12; introducing strong non-covalent interactions at key positions along the peptide will influence the unfolding pathway, inasmuch the order parameters presented here are concerned, overcoming the effect of other serious perturbations, such as the incorporation of a d-amino acid.

4.6.3 Concluding Remarks on d-Proline Studies

With electrostatics assuming once again the dominant role in conferring configurational endurance, simulations presented here support the view of gas-phase (mini-)protein conformation envisaging loss of folding co-operativity, which in solution is mediated by van der Waals interactions in conjunction with hydrophobic effects. Structural stability is manifested in long-lived and temperature-resistant local interactions, such as the networks of hydrogen bonds on which many secondary structures are based. Given an adequate amount of internal energy, the protein chain will seek to maximise strong intramolecular electrostatics, departing from the initial conformation; for Trp cage, the cost of losing the hydrophobic core is small in the gas-phase.

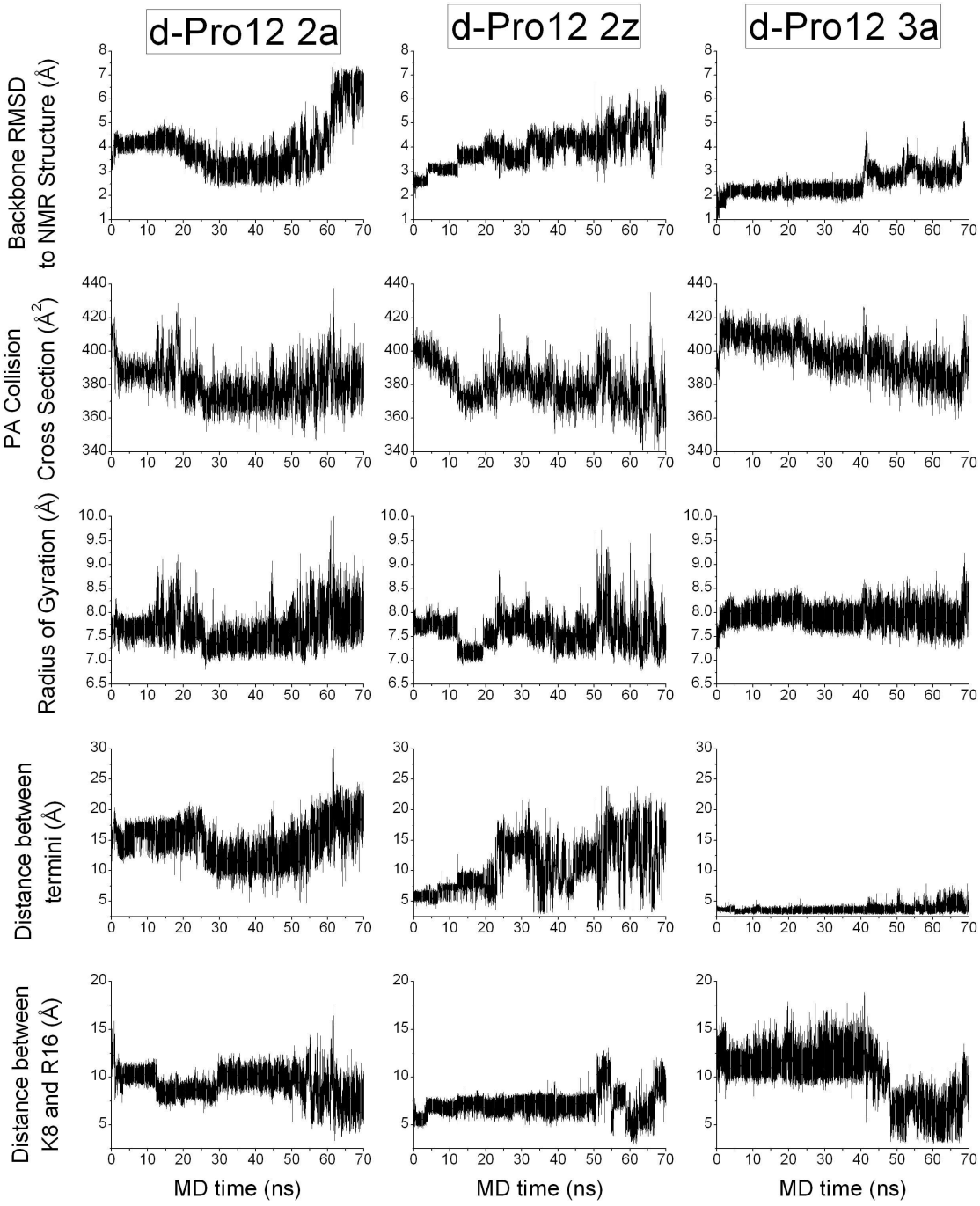


Figure 4.23
Analysis of MD trajectories for three charge assignments (from left to right: 2a, 2z, 3a) of d-Pro12. From top to bottom: Heavy-atom backbone RMSD to the NMR topology; Collision cross-sections calculated using the projection approximation method; Radii of gyration; distance between the N-terminal nitrogen atom and the C-terminal carbonyl carbon; distance between N_ζ of K8 and C_ζ of R16 calculated for geometries generated by MD simulations. Simulation temperature increments by 50K every 10ns of the trajectory: 0-10ns: 300K; 10-20ns: 350K; 20-30ns: 400K; 30-40ns: 450K; 40-50ns: 500K; 50-60ns: 550K; 60-70ns: 600K.

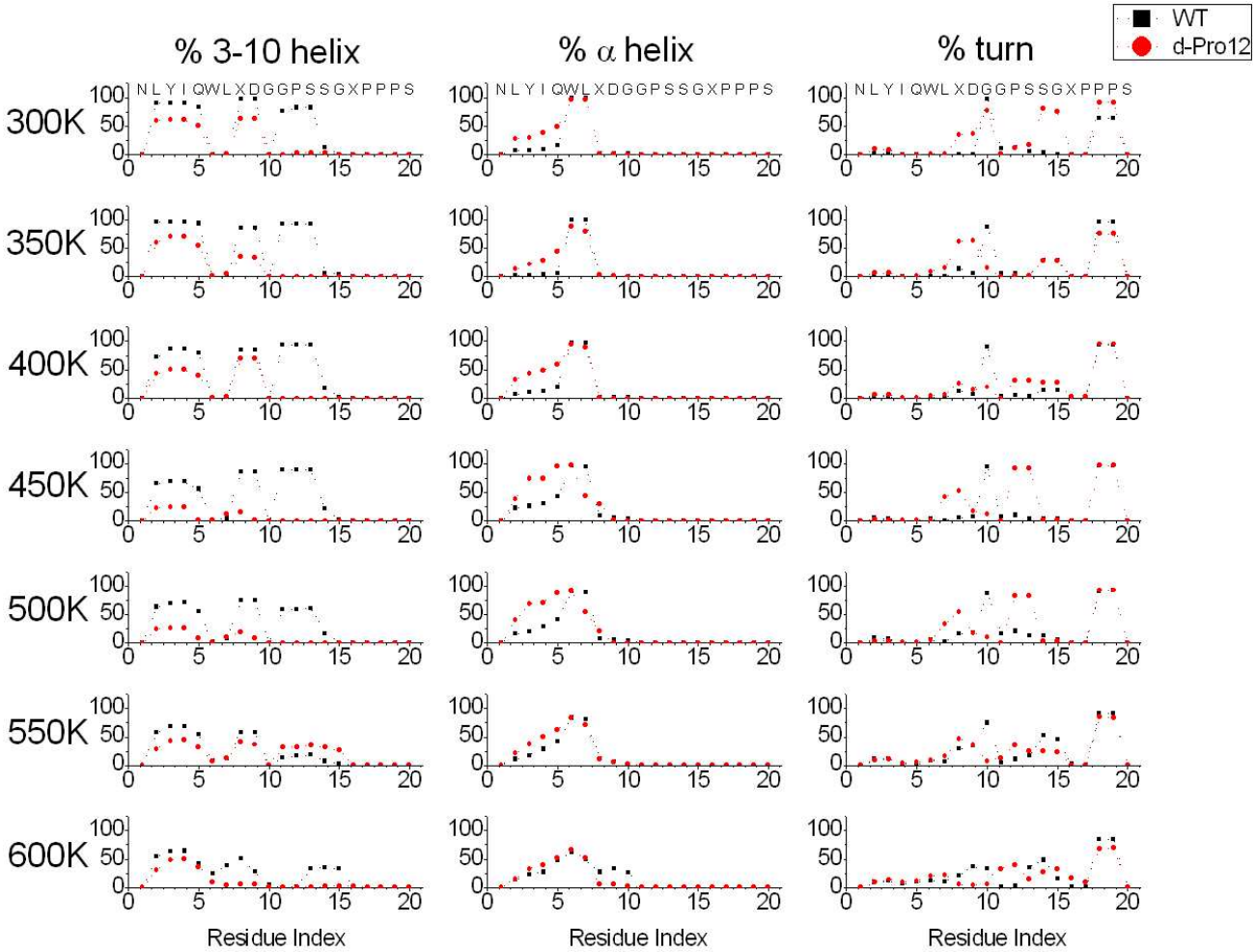


Figure 4.24
Average helical content and turn character for structures generated by MD at various temperatures of charge state 2a of WT (black squares) and d-Pro12 (red circles) calculated using the DSSP model. All constructs contain free termini. Averages are displayed as a function of peptide sequence. Left column, 3_{10} helical character; central column, α-helical character; right column, turn character.

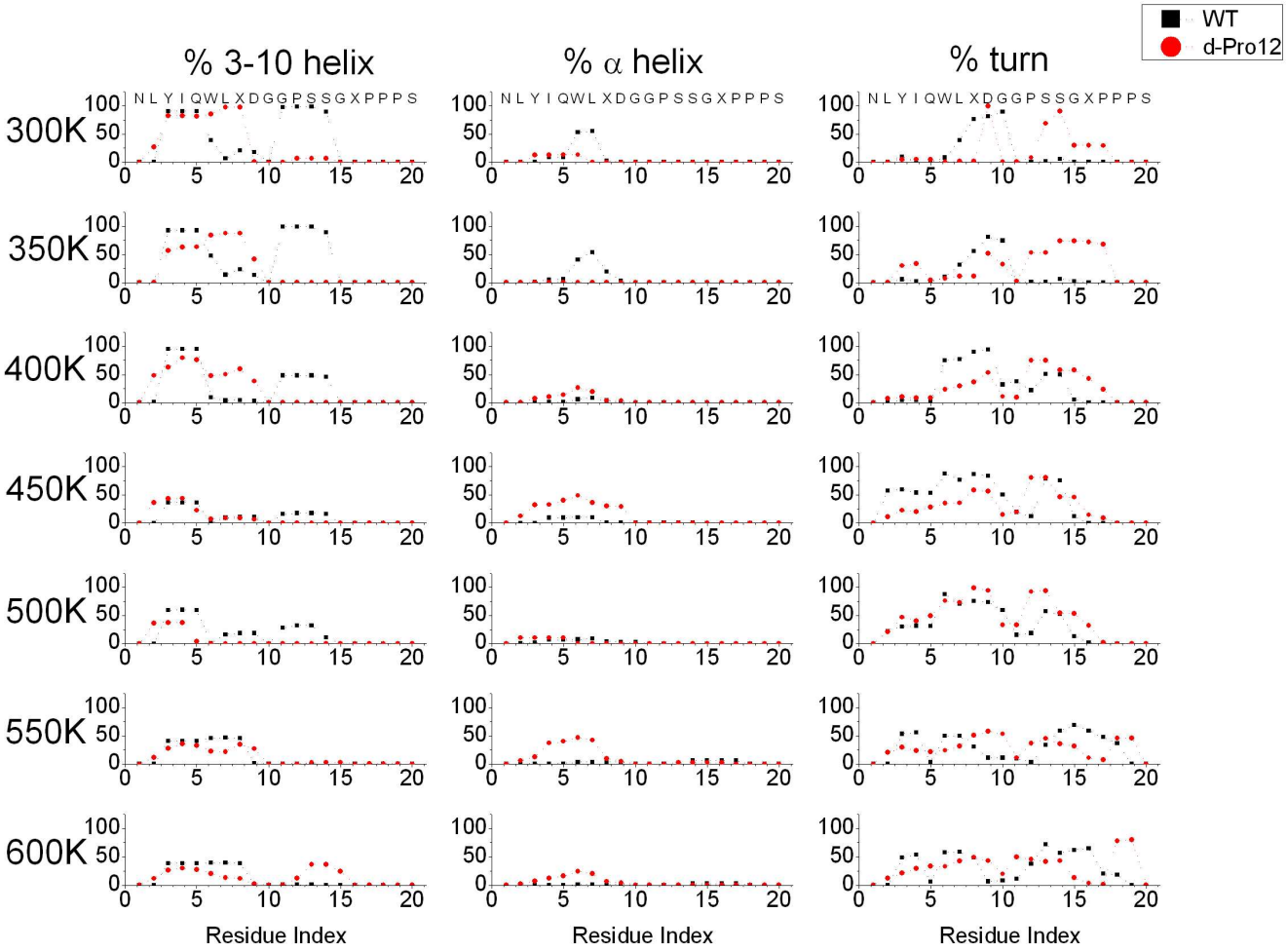


Figure 4.25
Average helical content and turn character for structures generated by MD at various temperatures of charge state 2z of WT (black squares) and d-Pro12 (red circles) calculated using the DSSP model. All constructs contain free termini. Averages are displayed as a function of peptide sequence. Left column, 3_{10} helical character; central column, α-helical character; right column, turn character.

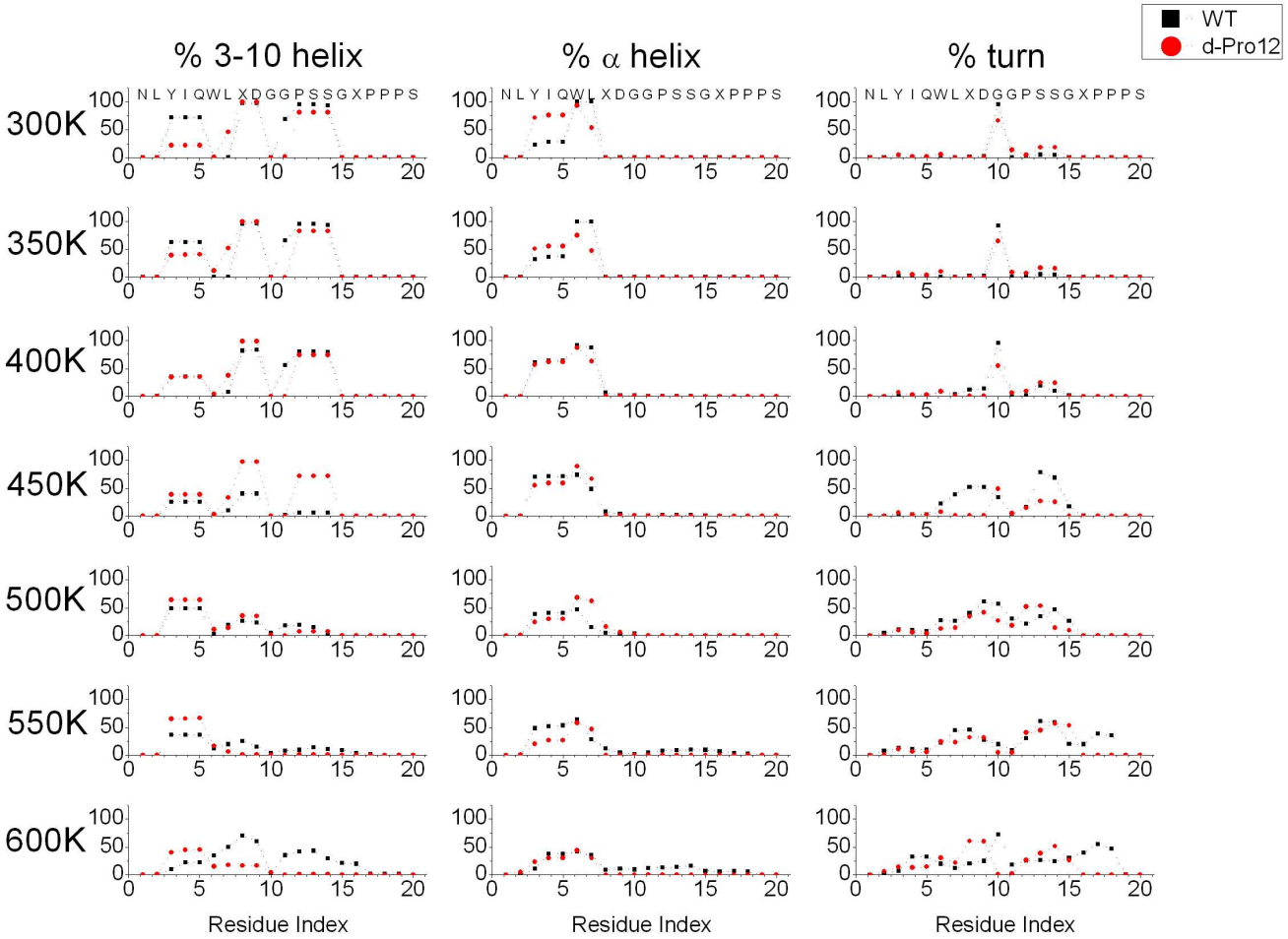


Figure 4.26
Average helical content and turn character for structures generated by MD at various temperatures of charge state 3a of WT (black squares) and d-Pro12 (red circles) calculated using the DSSP model. All constructs contain free termini. Averages are displayed as a function of peptide sequence. Left column, 310 helical character; central column, α-helical character; right column, turn character.

4.7 Summary

Early insights on the stability of Trp cage ions in the gas phase at energies higher than that of thermal energy at room temperature were deduced from measuring collisionally-induced dissociative parent-ion depletion (section 3.8.2). Arguably, these experiments monitor the stability of covalent structure (still, non-covalent structure has also been shown to have an effect on CID fragmentation[26]); however the message they convey is simple: addition of a proton reduces the energy required for fragmentation and diminishes the difference observed between peptide constructs with different side-chain proton affinities. Similarly, charge and charge location play a major role in determining the gas-phase conformation of desolvated Trp cage ions. High-temperature IMMS measurements and MD simulations have assisted to delineate the effect of charge and charge location, terminal capping groups (section 4.4), Arg-Lys and *vice versa* substitution (section 4.5) and a d-proline at position 12 (section 4.6) on the conformation of Trp cage ions.

CD measurements at increasing temperatures in the Far-UV and near-UV region (section 4.3) confirm that WT Trp cage is a protein with well-defined secondary and tertiary structure in 10mM NH_4CH_3COO which both contribute towards the stabilisation of the fold in a co-operative manner (as both unfolding curves give rise to a very similar transition temperature). In the absence of solvent the situation is different, as mobility data for d-Pro12 are similar to that for WT. As ions are studied at thermal energies by drift-tube experiments, stable tertiary structural elements may in fact form in the gas phase. Given the similarities between temperature-dependent collision cross sections of WT and d-Pro12, gas-phase geometries of these peptides may also be similar. Such similarities have been described for other d-amino acid containing Trp cage molecules in the gas phase[9, 10]. The main temperature-dependent trends of collision cross section in helium are shared between Trp cage molecules containing unprotected termini. Conversely, addition of protecting groups exerts a major effect. Viewed together with MD simulations, these findings imply that the presence of a free amine at the N-terminus strengthens the interaction between the peptide termini sufficiently to protect the gas-phase structure from unfolding.

As reviewed in section 4.2, the other main source of experimental information on the solvent-free structure of Trp cage ions at elevated temperatures is contained in the studies

published by Iavarone *et al.*[6-8]. In these FRET measurements in a Paul trap the authors report onset of unfolding at 360K for $[M+3H]^{3+}$ ions and 380K for $[M+2H]^{2+}$ ions of WT Trp cage ('unfolding' began at slightly lower temperatures for the D9N construct). Notably, Iavarone *et al.* measured quenching of intrinsic Trp fluorescence brought about by collisions between the indole group of Trp6 and the BODIPY-TMR dye linked covalently linked to the C-terminus of the peptide. The temperature range for an increase in cross section for $[M+3H]^{3+}$ ions (Figure 4.3) agrees very well to that observed for Trp quenching in the gas phase. However, $[M+2H]^{2+}$ ions remain compact up to 450K. This divergence can be explained in two ways, differences in experimental details notwithstanding:

- Lack of fluorescence quenching at low temperatures may be attributed to insufficient mobility of the BODIPY-TMR dye at low temperatures, rather than protection of the Trp6 side-chain for collisions
- Loss of packing around Trp6 may not be accompanied by a transition to more extended structures.

The latter hypothesis is supported by MD simulations presented in this chapter and by the fact that measured cross sections at low and intermediate temperatures lie far below that expected of the solution fold. Interestingly, thermally induced unfolding in the gas phase, for both $[M+2H]^{2+}$ and $[M+3H]^{3+}$ species, occurs at temperatures far greater than the ones measured by spectroscopic means in solution. Hence, once conformational rearrangements take place due to desolvation, the gas phase offers itself as an environment for isolating, separating and analysing macromolecular ions which give rise to complex non-covalent arrangements.

4.8 References

1. Pitera, J. W.; Swope, W., Understanding folding and design: Replica-exchange simulations of "Trp-cage" fly miniproteins. *Proceedings of the National Academy of Sciences of the United States of America* **2003,** *100* (13), 7587-7592.
2. Zhou, R. H., Trp-cage: Folding free energy landscape in explicit water. *Proceedings of the National Academy of Sciences of the United States of America* **2003,** *100* (23), 13280-13285.
3. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995). *Journal of the American Chemical Society* **1996,** *118* (9), 2309-2309.

4. Hudaky, P.; Straner, P.; Farkas, V.; Varadi, G.; Toth, G.; Perczel, A., Cooperation between a salt bridge and the hydrophobic core triggers fold stabilization in a Trp-cage miniprotein. *Biochemistry* **2008,** *47* (3), 1007-1016.
5. Iavarone, A. T.; Paech, K.; Williams, E. R., Effects of charge state and cationizing agent on the electron capture dissociation of a peptide. *Analytical Chemistry* **2004,** *76* (8), 2231-2238.
6. Iavarone, A. T.; Parks, J. H., Conformational change in unsolvated Trp-cage protein probed by fluorescence. *Journal of the American Chemical Society* **2005,** *127* (24), 8606-8607.
7. Iavarone, A. T.; Duft, D.; Parks, J. H., Shedding light on biomolecule conformational dynamics using fluorescence measurements of trapped ions. *Journal of Physical Chemistry A* **2006,** *110* (47), 12714-12727.
8. Iavarone, A. T.; Patriksson, A.; van der Spoel, D.; Parks, J. H., Fluorescence probe of Trp-cage protein conformation in solution and in gas phase. *Journal of the American Chemical Society* **2007,** *129* (21), 6726-6735.
9. Patriksson, A.; Adams, C.; Kjeldsen, F.; Raber, J.; van der Spoel, D.; Zubarev, R. A., Prediction of N-C-alpha bond cleavage frequencies in electron capture dissociation of Trp-cage dications by force-field molecular dynamics simulations. *International Journal of Mass Spectrometry* **2006,** *248* (3), 124-135.
10. Patriksson, A.; Adams, C. M.; Kjeldsen, F.; Zubarev, R. A.; van der Spoel, D., A direct comparison of protein structure in the gas and solution phase: The TRP-cage. *Journal of Physical Chemistry B* **2007,** *111*, 13147-13150.
11. Adams, C. M.; Kjeldsen, F.; Zubarev, R. A.; Budnik, B. A.; Haselmann, K. F., Electron capture dissociation distinguishes a single D-amino acid in a protein and probes the tertiary structure. *Journal of the American Society for Mass Spectrometry* **2004,** *15* (7), 1087-1098.
12. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., ELECTROSPRAY IONIZATION-PRINCIPLES AND PRACTICE. *Mass Spectrometry Reviews* **1990,** *9* (1), 37-70.
13. Paschek, D.; Hempel, S.; Garcia, A. E., Computing the stability diagram Trp-cage miniprotein of the. *Proceedings of the National Academy of Sciences of the United States of America* **2008,** *105* (46), 17754-17759.
14. Ahmed, Z.; Beta, I. A.; Mikhonin, A. V.; Asher, S. A., UV-resonance Raman thermal unfolding study of Trp-cage shows that it is not a simple two-state miniprotein. *Journal of the American Chemical Society* **2005,** *127* (31), 10943-10950.
15. Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F., NAKED PROTEIN CONFORMATIONS - CYTOCHROME-C IN THE GAS-PHASE. *Journal of the American Chemical Society* **1995,** *117* (40), 10141-10142; Clemmer, D. E.; Jarrold, M. F., Ion mobility measurements and their applications to clusters and biomolecules. *Journal of Mass Spectrometry* **1997,** *32* (6), 577-592; Hoaglund-Hyzer, C. S.; Counterman, A. E.; Clemmer, D. E., Anhydrous protein ions. *Chemical Reviews* **1999,** *99* (10), 3037-3079.
16. Sobott, F.; Hernandez, H.; McCammon, M. G.; Tito, M. A.; Robinson, C. V., A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Analytical Chemistry* **2002,** *74* (6), 1402-1407.
17. Lin, M. X.; Ahmed, Z.; Taormina, C. R.; Somayajula, K. V., A quadrupole/time-of-flight mass spectrometry study of Trp-cage's conformation. *Journal of the American Society for Mass Spectrometry* **2007,** *18* (2), 195-200.
18. Dobo, A.; Kaltashov, I. A., Detection of multiple protein conformational ensembles in solution via deconvolution of charge-state distributions in ESI MS. *Analytical Chemistry* **2001,** *73* (20), 4763-4773.
19. Wagner, D. S.; Anderegg, R. J., CONFORMATION OR CYTOCHROME-C STUDIED BY DEUTERIUM-EXCHANGE ELECTROSPRAY-IONIZATION MASS-SPECTROMETRY. *Analytical Chemistry* **1994,** *66* (5), 706-711.
20. Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H., Designing a 20-residue protein. *Nature Structural Biology* **2002,** *9* (6), 425-430.
21. Harrison, A. G., The gas-phase basicities and proton affinities of amino acids and peptides. *Mass Spectrometry Reviews* **1997,** *16* (4), 201-217.
22. Hudgins, R. R.; Ratner, M. A.; Jarrold, M. F., Design of helices that are stable in vacuo. *Journal of the American Chemical Society* **1998,** *120* (49), 12974-12975.
23. Stearns, J. A.; Boyarkin, O. V.; Rizzo, T. R., Spectroscopic signatures of gas-phase helices: Ac-Phe-(Ala)(5)-Lys-H+ and Ac-Phe-(Ala)(10)-Lys-H+. *Journal of the American Chemical Society* **2007,** *129*, 13820-+; Stearns, J. A.; Seaiby, C.; Boyarkin, O. V.; Rizzo, T. R., Spectroscopy and conformational preferences of gas-phase helices. *Physical Chemistry Chemical Physics* **2009,** *11* (1), 125-132.

24. Bush, M. F.; Forbes, M. W.; Jockusch, R. A.; Oomens, J.; Polfer, N. C.; Saykally, R. J.; Williams, E. R., Infrared spectroscopy of cationized lysine and epsilon-N-methyllysine in the gas phase: Effects of alkali-metal ion size and proton affinity on zwitterion stability. *Journal of Physical Chemistry A* **2007,** *111* (32), 7753-7760; Forbes, M. W.; Bush, M. F.; Polfer, N. C.; Oomens, J.; Dunbar, R. C.; Williams, E. R.; Jockusch, R. A., Infrared spectroscopy of arginine cation complexes: Direct observation of gas-phase zwitterions. *Journal of Physical Chemistry A* **2007,** *111*, 11759-11770; Dunbar, R. C.; Steill, J. D.; Polfer, N. C.; Oomens, J., Peptide Length, Steric Effects, and Ion Solvation Govern Zwitterion Stabilization in Barium-Chelated Di- and Tripeptides. *Journal of Physical Chemistry B* **2009,** *113* (31), 10552-10554.
25. Kabsch, W.; Sander, C., DICTIONARY OF PROTEIN SECONDARY STRUCTURE - PATTERN-RECOGNITION OF HYDROGEN-BONDED AND GEOMETRICAL FEATURES. *Biopolymers* **1983,** *22* (12), 2577-2637.
26. Tsaprailis, G.; Nair, H.; Somogyi, A.; Wysocki, V. H.; Zhong, W. Q.; Futrell, J. H.; Summerfield, S. G.; Gaskell, S. J., Influence of secondary structure on the fragmentation of protonated peptides. *Journal of the American Chemical Society* **1999,** *121* (22), 5142-5154.

5

Gas-phase Geometries of Oligomeric Aggregates of the Amyloidogenic Transthyretin Fragment 105-115

5.1 Abstract

Amyloid formation is a result of often pathological polypeptide aggregation, but also has potential for achieving control over molecular architectures at a molecular level – with an immediate effect on macroscopic properties. Transthyretin fragment 105-115 (YTIAALLSPYS; derived from human transthyretin, the major component of amyloid plaques from familial amyloid polyneuropathic patients) is an amyloidogenic endecapeptide which, due to its convenient size, can serve as an amyloid model. The early aggregation states of TTR105-115 have been probed by MS and IMMS. Models of early peptide aggregates' topology have been interrogated using molecular mechanics methods. Two classes of conformations were examined: antiparallel β-sheets and globular geometries. Collision cross sections of the models were compared to the experimental data in order to determine the configuration of TTR105-115 oligomeric assemblies.

5.1.1 Health, Disease and Technology: the Many Facets of Amyloid

Tissue isolates from a wide range of diseased states have been found to contain corpuscules proximal to affected sites. Virchow characterised deposits obtained from autopsies of deceased patients in the mid-19th century and observed a positive iodine staining, characteristic of starch; he suggested the term amyloid for those pathological tissue formations as he considered them analogous to plant cellulose (Virchow (1853) as cited by Puchtler and Sweat[1]). In 1859, Friedrich and Kekulé (Puchtler and Sweat, *ut supra*[1]) reported the presence of nitrogen in amyloid plaques thus sowing the first seed of what is nowadays the most dominant view of amyloid: a protein aggregate. More than 150 years separating us from this earliest amyloid research having lapsed, it is now known that amyloid deposits are composed of a proteinaceous fibrillar core, with the following characteristics:

- Each type of fibril is associated with a single protein, it being its major component[*]
- Fibrils occupy a "cross-β" structure; constituent peptides self assemble into extended (often intertwined) β-sheets with the peptide axis perpendicular to the fibre axis.

Several physical properties are conferred to amyloid by virtue of its very regular structure and which, moreover, contribute to its identification. Amyloid fibrils bind to the diazo dye Congo Red, causing a red-shift of its absorbance peak and an apple-green birefringence (an assay put forward by Bennhold in 1922, as cited by Sipe[2]); amyloid also enhances the fluorescence of thioflavin T[2].

Diseases associated with amyloid formation (collectively termed amyloidoses) are diverse, causing degeneration of particular tissues, can be hereditary, sporadic or even infective, and, despite macroscopic similarities of amyloid, constituent proteins are not evolutionarily or structurally (in the strictest notion) related. They include Alzheimer's disease (caused by Alzheimer's beta (Aβ) peptide), Parkinson's syndrome (involving α-

[*] It must be noted that precludes heterogeneous sequences from forming amyloid fibres *in vitro*.

synuclein), Huntington's disease (huntingtin), type-II diabetes (islet amyloid polypeptide, IAPP), dialysis-related amyloidosis (β2-microglobulin), cataracts (γ-crystallin), transmissible spongiform encephalopathies (such as Creutzfeld-Jacobs syndrome, involving prion proteins, PrP), spinocerebellar ataxiae, familial amyloidotic polyneuropathy (FAP, caused by TTR and its many mutants) and senile systemic amyloidosis (also TTR)[2, 3]. Amyloidoses are essentially incurable to date and their severity has fuelled research which has begun to reveal the basic processes underlying them. However, amyloid-like structures are not only encountered in diseased states. Several examples of biologically functional amyloid are known to date[4], including surface adhesion of some bacteria, epigenetic factors in yeast[5], melanin polymerisation in mammalian skin cells[6]; even the spider dragline silk and other insect silks are composed of structures reminiscent of amyloid[7]. The parallel between spider silk and amyloid is most intriguing since the former has been " hailed as a 'super-fibre' "[8] with a tensile strength 1.1GPa, an initial Young's modulus of 8GPa and a breaking elongation of 40% (values reported here for *Nephila edulis* dragline silk). Corresponding properties of amyloid fibrils are comparable: using AFM techniques Smith *et al.*[9] found bovine insulin amyloid fibrils to have a tensile strength of 0.6GPa and a Young's modulus of 3.3GPa. The robustness of amyloid fibrils along with their ability to self-assemble in a controlled fashion under regulated conditions into an ordered structure not only explains their persistence in diseased states and the utilisation of similar structures by living organisms, but creates the possibility of designing novel nanomaterials exploiting these properties; this direction has already been pursued to some extent (as reviewed recently by Channon and MacPhee[10]). Among the potential applications for amyloidogenic peptides may exploit their ability to self-assemble in order to scaffold the formation of potentially useful nanomaterials: in an intriguing study, Reches and Gazit employed hollow nanotubule-forming diphenylalanine[*] moieties to cast silver nanowires of a defined length[11]; notably these examples leave aside all possible biotechnological and medical applications. Furthermore, amyloidogenic peptides need not be based on natural sequences alone. Nowadays a biophysical chemist's repertoire includes design strategies

[*] A sequence motif found in the Alzheimer's syndrome-associated Aβ peptide

for peptides aggregating into amyloid-like fibrils and higher order structures, with additional desirable (and adjustable) properties; such efforts have been collectively described by Pastor *et al.*[12]. Clearly, a greater understanding of the mechanism of amyloid fibrillogenesis may not only illuminate the origins of pathogenicity in amyloid diseases but will benefit many sectors of academic research, from comprehending the mechanism of "infective" protein conformations of prions to the design and controlled formation of robust self-assembling nanoscale structures.

5.2.1 Transthyretin: Structure and Implications in Disease

Transthyretin (TTR)[*] is a 127-amino acid plasma protein, expressed predominantly in the liver, choroid plexus and retina; it acts as a major transporter of thyroxine (and other thyroid hormones) and retinol (vitamin A)[13]. TTR is abundant in blood serum with a concentration of ca. 0.2-0.4mg/ml, its concentration in cerebrospinal fluid being 20- to 50-fold less[13]. The tetramer can be conceptualised as being a dimer of dimers. The extensive β-strand content and the presence of intermolecular β contacts may be very significant for explaining the fibrillogenic properties of dissociated monomers. Several diseased states are manifestations of TTR aggregation into amyloid: FAP (one of the most abundant inherited polyneuropathies), familial amyloidotic cardiomyopathy (FAC), and senile systemic amyloidosis (SSA, affecting >20% of people above 90 years of age)[14]. Due to the severity of amyloidoses mutant TTRs have been a subject of intensive investigation leading to several variants even being crystallised; defying expectations, the resulting tetrameric structures do not reveal significant differences[15], perhaps with the exception of the aggressively amyloidotic L55P variant (which, again, adopts an overall backbone topology nearly identical to the wild-type protein). However, it must be noted that the highly ordered and congested environment of a crystal limits scope of crystallographic techniques for investigating highly dynamic phenomena. Furthermore, albeit FAP being produced by single-point mutations, SSA occurs in individuals homozygous for the wild-type sequence[14, 16] which is also amyloidogenic *in vitro* under mildly denaturing conditions[17]. The composition of *ex vivo* fibrils is very heterogeneous,

[*] Also called prealbumin, thus named after its property of migrating less far than albumin on an SDS-PAGE gel.

full-length TTR, mutant and wild-type, being overwhelmed by a range of fragments – presumably produced proteolytically[18]. Destabilisation of the native tetrameric fold seems to be the first common step for fibrillogenesis of all TTR variants[19, 20]. Evidently, a more involved model of pathological amyloidogenesis would be necessary to account for the whole scope of phenomena that have been observed. A reductionist approach would deem desirable to engineer smaller systems in order to capture the key steps of amyloid formation.

5.2.2 Some Transthyretin Fragments are Fibrillogenic

Usage of smaller oligopeptide models offers a number of advantages: the relative ease of synthesising them at reasonably high yields allows the correlation between structure and composition to be studied more systematically; simulations of minimal amyloidogenic systems also require fewer resources making a more thorough treatment possible. In 1990 Gustavsson *et al.*[21, 22] reported the synthesis of a range of TTR fragments which they subsequently tested for their ability to generate amyloid fibrils *in vitro*. They discovered that TTR10-20 (CPLMVKVLDAV) and TTR105-115 (YTIAALLSPYS), corresponding to β-strands A and G respectively, both proximal to the hydrophobic thyroxine binding cleft, readily formed amyloid-like fibrils whereas TTR50-60 (SESGELHGLTT), in which only two amino acids are arranged in a formally β-strand conformation in the crystal structure, did not form fibrils under the conditions tested. Additionally, TTR10-20 self-associated rapidly in acidified solutions, much like the full-length protein, whereas TTR105-115 was fibrillogenic even at neutral pH. Due to the association of TTR with diseased states and the convenience of using smaller peptides as amyloid models, the identification of readily amyloidogenic TTR fragments triggered further biophysical investigations, whose description also serves as an overview of amyloid fibril structure.

5.2.3 TTR105-115 Amyloid Fibril Structure

Soon after the fibrillogenic character of TTR fragments was reported, the fibres were subjected to characterisation by X-ray diffraction[22, 23]. The X-ray diffraction patterns were typical of those displayed by amyloid fibrils, showing two dominant reflections: 1)

a 4.7Å meridian reflection (perpendicular to the fibre axis) which is assigned to the spacings of the hydrogen-bonded β-strands and 2) a 8-10Å equatorial reflection (parallel to the fibre axis) which corresponds to the separation of the stacked β-sheets[24]. Compared to the other constructs examined, TTR105-115 displayed further reflections which indicated an ordered structure. Indeed, the structure of TTR monomers in the fibrillar form was solved using solid-phase NMR techniques and isotopic labelling[25-27]; these studies revealed that individual TTR105-115 molecules adopt an extended β-strand conformation. Moreover, even the peptide side-chains (with the exception of Y105 and S115) were characterised by a great degree of configurational order allowing them to be compared to microcrystalline rather than disordered states[27]. The first studies on TTR fragments showed that TTR10-20 and TTR105-115 assembled *in vitro* into fibrils with a diameter of ~10nm, being significantly smaller than fibrils formed by the full-length protein[21, 22]. Later more detailed morphological studies using AFM and EM found TTR105-115 fibrils incubated in acidified aqueous solutions containing 10% ACN for two weeks had a diameter in the 7-12nm range and were about 1μm in length[28]. Individual fibrils imaged with AFM also displayed a pleated, rope-like appearance supporting the hypothesis stating that "mature" fibrils are constructed from at least two intertwined protofilaments, each protofilament containing several stacked β-sheets[28] (in an extended conformation, TTR105-115 has a length of 4nm). A certain degree of polymorphism on fibril structure – regarding especially the number of protofilaments per fibre – is to be expected and has been witnessed in other systems, e.g. fibrils formed from the immunoglobulin light chain are thought to contain either four or six protofilaments[29]. By depositing insulin fibres on patterned surfaces and using the AFM tip to apply forces of pN magnitude, Smith *et al.* measured the fibrils' rigidity, finding them surprisingly stable[9]. Comparing the appearance of insulin and TTR105-115 fibrils as depicted by AFM, Mesquida *et al.* concluded that TTR105-115 fibrils are even stiffer, being less curved[30]. Furthermore, application of hydrostatic pressure of the order of GPa on mature TTR105-115 fibrils does not impart any major perturbations of the resulting FTIR spectrum[31]. Overall, TTR105-115 forms mechanically and chemically robust fibrils, characterised by notable order at the molecular level. Being a relatively small

synthetic peptide, it favours its use as a model for studying amyloid but also has the properties which render it suitable for use in novel applications. Recently Gras *et al.* have modified TTR105-115 by the C-terminal addition of the RGD sequence to optimise its ability to form a template for cell adhesion for tissue engineering applications[32].

5.3.1 Cross-β Structure May be a Generic Stable Conformation of Protein Aggregates

What proteins are capable of forming amyloid fibrils? More than twenty proteins form amyloid fibrils associated with human disease all of which are structurally diverse and none are evolutionarily related[2, 3]. The structural variation of "parent" sequences, the ability to control self-assembly of small peptides to the extent of formulating criteria for design of amyloidogenic sequences[3] the common overall configuration of amyloid fibrils as well as their physiochemical stability imply that the cross-β structure may be in many cases the global potential energy minimum of large protein aggregates[33], native protein folds being essentially metastable. These theories are corroborated by the discovery of fibrillogenic potential in proteins not linked to amyloid formation in a biological context, upon mild denaturation: examples include bovine PI3-kinase SH3 domain, acylphosphatase and the *E. coli* HypF protein[3]. The self-assembly of even very small peptides (2-3 amino acids) into ordered fibrillar geometries is considered so fundamental that it has even been utilised to forming an origin-of-life hypothesis[34]. Due to their thermodynamic stability[35], it may be possible to form amyloid-like fibrils out of virtually any sequence, thus overriding local conformational preferences. This phenomenon is demonstrated even by the TTR105-115 peptide studied herein, whereby P113 adopts a β-conformation in the fibrillar state[26, 27].

5.3.2 Non-fibrillar Aggregates Elicit a Cytotoxic Effect

Amyloid deposits are most famous due to their association with disease, yet the pathogenic mechanism of amyloidoses remains elusive with the role of amyloid fibrils in causing the diseased phenotypes being debated. Conditions caused by necrosis of particular tissues, such as FAP, AD, PD and other neurodegenerative amyloidoses, may develop from the cytotoxic character of protein deposits. Despite deposition of fibrillar aggregates being a histological hallmark of amyloidosis, a straightforward relationship between appearance of fibrillar deposits and disease onset (in both human subjects and laboratory animals) has not been obtained, leading to the conclusion that fibrils are not the primary causative factor but a side-effect of disease progression[36]. Furthermore, numerous studies have identified non-fibrillar aggregates as toxic to cultured cells showing simultaneously that mature fibrils are physiologically inert, which agrees with their mechanical and chemical robustness[36]. Early neurodegeneration in FAP patients takes place in the presence of non-fibrillar TTR, as evidenced by anti-TTR and Congo Red staining of tissue sections[37]. Additionally, Reixach *et al.*[38] found non-native pre-fibrillar native and mutant TTR monomers and oligomers formed *in vitro* to be cytotoxic when tested against human neuroblastoma cells. According to both studies, fibrillar TTR aggregates are not significantly cytotoxic. Intriguingly, amorphous aggregates originating from non-disease-associated proteins also display cytotoxic traits: Bucciantini *et al.*[39] detected assemblies occurring prior to fibrillation of bovine PI3-kinase SH3 domain and *E. coli* HypF protein able to kill mouse fibroblasts and pheochromocytoma cells, a trait not exhibited by mature fibres. These observations indicate that aggregating aberrantly folded proteins administered extracellularly can be lethal to cells, leading to the hypothesis that amyloid fibril formation may act as an intrinsic protective mechanism against the toxic effects of non-fibrillar assemblies; such considerations can consult the pursuit of therapeutic strategies.

5.4 Fibrillogenesis Mechanism

Amyloid formation *in vitro* generally follows a nucleation-and-growth mechanism. Fibrils appear only after an initial lag phase, after which mature fibrils appear rapidly

and lengthen in an exponential manner, incorporating most soluble peptide molecules. Addition of preformed fibrils in non-fibrillar solutions induces fibril formation, shortening the lag phase ("seeding"). These features of the fibrillation mechanism indicate that during the initial lag phase peptides self-associate and rearrange into a critical nucleus capable of supporting growth of the first fibril which subsequently seeds further fibril growth. In the absence of further environmental perturbations, fibril formation is irreversible.

The details of amyloid formation have been the subject of intensive study over the last two decades and a wide range of techniques has been employed towards this aim.

Amyloid formation can be conceptualised as comprising of three stages:

- Nucleation
- Fibril growth
- Fibril maturation

Major structural changes occur during each of the above conceptual steps; we shall now proceed with describing each stage separately, illustrating individual events constituting fibril formation with particular examples. The Amyloid-β peptide associated with AD will be the most oft-quoted example since it is the most studied; yet we will refer to findings based on observations of TTR and TTR fragment fibrillogenesis.

5.4.1 Events Leading to Fibril Nucleation

Significant conformational rearrangements need to take place for the conversion of the "native" configurational ensemble to a cross-β architecture characteristic of amyloid. Thus amyloidogenic polypeptides are expected to undergo a conformational change into an aggregation-prone geometry. Exposure of hydrophobic regions is known to underlie aggregation in aqueous environments. Numerous amyloidogenic peptides display "random coil" behaviour (i.e. their structure must be represented as an equilibrium between a number of different geometries): such systems include Aβ, α-synuclein and the vast majority of short fibrillogenic peptide fragments, designed or natural. The aggregation tendencies of smaller or rheomorphic polypeptides shouldn't be striking since such systems present hydrophobic regions at their solvent-accessible surface.

Proteins that can be portrayed by a single native fold require its destabilisation for amyloidogenesis to take place. Several pathways can contribute towards generation of aggregating structures *in vivo*. Topical environmental changes may alter protein stability; aberrant peptide processing and/or clearance may lead to local accumulation of aggregating species; finally, mutations may destabilise the native fold. The latter is often invoked as the most likely cause of FAP[40]. The infringement of amyloidogenic mutations on the TTR tetramer formation suggest native-state stabilisation (by either allosteric or competitive ligand binding or supply of tetramer-stabiliding TTR mutants) as a viable therapeutic strategy for FAP[41, 42]. Fibrillogenesis, especially at its early stages, is a highly dynamic process and can only be poorly represented by steady-state models.

In line with this view, soluble amyloidogenic monomers do not remain latent until the onset of amyloid fibril formation. A wide range of structures, with characteristic morphologies, are witnessed during the early stages of fibrillogenesis and can be broadly classified into three families:

- Soluble oligomers
- Large amorphous, flocculous aggregates
- Protofibrils

Protofibrils[*], which often co-exist with larger aggregates[43-46], are structurally heterogeneous as they can comprise of small assemblies appearing as spheres or ellipsoids but are also manifested as flexible worm-like fibrils (as opposed to mature fibrils which are relatively rigid and straight)[19, 43-45]. Spherical/ellipsoid aggregates measure a few nanometres in diameter[47] (e.g. ~5nm or ~8nm in the case of full-length TTR[19]) suggesting that they are oligomeric. Protofibrils display further polymorphisms: Cardoso *et al.*[19] have detected both thin (4nm) and thicker (8nm) filaments occurring in aggregating solutions of wild-type TTR, thicker filaments occasionally displaying a left-handed helical twist with 17nm periodicity. Furthermore, the existence of annular structures (apparently formed by circularisation of flexible protofibrils) has been reported for Aβ and α-synuclein[36]; it has even been suggested that cytotoxicity of

[*] Different from protofilaments, the core structures of mature fibres

protofibrillar structures may be effected by formation of unregulated pores on cell membranes[36]. Only spherical/ellipsoid aggregates are present during the very early steps of peptide aggregation, but short protofibrils soon make their appearance[44, 45]. Protofibrils lengthen and reduce in number the longer aggregating samples are incubated before mature fibril formation[44, 45]. The beads-on-string appearance of larger protofibrils invites the speculation of them being assembled from the spherical assemblies and possibly further stabilised by a conformational change. Protofibril growth rate and morphology is responsive to environmental conditions, such as pH, temperature and ionic strength[43].

When mature fibrils make their appearance protofibrils reduce in number[19, 43-45], indicating that mature fibril growth depletes the protofibril peptide pool. Since protofibrils are in equilibrium with oligomeric states[43-45] whether they are 'on-pathway' or 'off-pathway' to mature fibril formation remains unclear. Nevertheless, a number of hypotheses can be accommodated: 1) protofibrils of a particular size undergo a conformational change which nucleates fibril growth 2) the fibril nucleating structure is formed within protofibrils or larger assemblies but must be subsequently released for mature fibril growth to ensue by monomer addition 3) the fibril nucleus is structurally altogether different from protofibrils, the latter being a kinetic trap of fibril generation.

The aggregation mechanism of different amyloidogenic peptides differ significantly, as do the details of the appearance of the produced fibrils; however the numerous similarities are noteworthy. Protofibrils do appear in aggregating mixtures of a number of unrelated proteins (e.g. TTR, Aβ, α-synuclein and PI3-SH3 domain), all displaying similar appearance when examined microscopically[3, 19, 36, 43-45, 48]. Kayed *et al.*[49] have raised antibodies against Aβ early aggregates in rabbits which selectively recognised protofibrils over smaller soluble aggregates and mature fibrils; intriguingly these antibodies also bound to protofibrils from other amyloidogenic proteins. Furthermore, when supplied in excess, the antibodies significantly compromised the *in vitro* cytotoxicity of protofibrils. Such findings suggest that protofibrils may in fact share certain structural elements, transcending primary sequence variation amongst aggregating polypeptides.

Amorphous aggregates occurring very rapidly may result from non-specific aggregation of partially soluble peptides, and have been evidenced e.g. for Aβ by fluorescence correlation spectroscopy[46] and EM[50], appearing as a large diffuse peptide assembly. However, examination of such assemblies display structural features at the edges of large peptide aggregates which resemble protofibrils[19, 43, 48, 51] thus the hypothesis that amorphous aggregates are in fact assemblies of protofibrils is also justifiable.

Logically, the formation of oligomeric species is necessary for the generation of larger aggregates. Due to their small size and transient nature, oligomeric states are the most cumbersome to follow with most biophysical techniques. FCS has been successful in detecting oligomeric aggregates but does not differentiate between individual oligomers[46]. Separation techniques such as electrophoretic techniques, size-exclusion chromatography and analytical ultracentrifugation require that species are not interconverting during experimental timescales to render them clearly distinguishable. Insights into the early oligomerisation states of Aβ can be afforded however by chemical[52] and photochemical[53, 54] cross-linking techniques.

Of particular relevance to the work presented herein is the appearance of oligomer distributions. Fibrillogenic peptides (Aβ, TTR, calcitonin) normally have narrower distributions but at the same time centred at larger aggregates (trimer, tetramer) than non-fibrillogenic peptides (pituitary adenylate cyclase activating polypeptide and growth hormone releasing factor), which are characterised by broad distributions but with the monomeric or dimeric state being most abundant. Distributions of amyloidogenic peptides are often non-gaussian[*]. Both experimental evidence and empirical theoretical

[*] A brief discussion of the multimer distributions in aggregating species may illuminate this point. Assuming negligible activation energy barriers to aggregation, the relation between thermal energy $k_B T$ and energy gain due to self-association will determine the resulting distribution at equilibrium. At either low concentrations or when thermal enegy is high compared to aggregation energy gain, monomers will be the dominant species and the propensity of higher oligomers will drop monotonously. When the aggregation energy gain is higher but comparable to thermal energy, all aggregation states will have similar (albeit not equal) probabilities. Large aggregates will prevail at sufficiently high concentrations when the aggregation energy is significantly larger than that of thermal energy and a sufficiently high aggregation energy gain will lead to the solute separating from the solvent. Starting from an all-monomeric distribution, systems of the latter size will gradually occupy higher oligomerisation states, at rates dependent among others on their diffusion properties. Assuming a peptide could exist in (at least) two conformational ensembles with different aggregation propensities, a distribution of multimers decreasing with aggregate size overlapping with a peak at higher multimeric states could result. Needless

models showed that increasing the cross-linking (irradiation) time resulted into broader distributions with the distribution being centred at higher-order oligomers. These findings suggest that properties of oligomer distributions may be unique of self-associating peptides and may even yield insights into the mechanism of the initial self-association events. Mass spectrometry can also differentiate between discreet oligomerisation states and has been employed in the study of amyloidogenesis; such research will be reviewed in section 5.5.

From the limited literature cited above it is clear that multiple events precede fibril formation. Ideally monomeric peptides rapidly establish an equilibrium with oligomeric aggregates which possibly form an oligomeric nucleus for protofibril formation; protofibrils lengthen by either addition of monomers, oligomers or end-to-end annealing of shorter protofibrils and may even self-associate to form large assemblies[19, 43, 48, 55]. This series of self-association (and dissociation) events is accompanied by significant conformational changes. For example, aggregating mixtures of Aβ monitored by CD over weeks have revealed the presence of a transient helical intermediate, albeit the final product (fibrils) being a β-strand-based structure[56, 57]. Nevertheless, Aβ protofibrils fractionated by SEC display Congo Red, ThT binding and a CD trace characteristic of β sheets indicating that the α-helical structure must be attributed to a different peptide assembly[57], yet whether it is soluble oligomers or large assemblies that adopt the helical structure is left to speculation. Aggregating mixtures thus contain a variety of not only quaternary, but also secondary and tertiary structures whose exact nature and dynamics remain to be elucidated. Eventually this process forms a structure nucleating mature fibril formation.

5.4.2 Fibril Growth

The depletion of other peptide aggregates into fibres with characteristic β-strand structure alone unifies all amyloid fibrillogenic reactions irrespective of primary sequence. Nevertheless, fibrils, like their precursors, are not devoid of heterogeneity.

---

to say that the conformational flexibility of polypeptides diminishes the predictive powers of such simple considerations yet can still serve to illustrate the general principles: specific conformational preferences may well give rise to unusual and complex oligomer distributions.

The morphology of the final aggregation product can be influenced by solution conditions *in vitro*. *In vivo* fibrils are heterogeneous in both composition and morphology: to illustrate this point, *ex vivo* amyloid deposits of TTR contain both a gamut of TTR fragments in addition to tissue-specific factors and amyloid P[13]; additionally, fibrillar deposits from FAP patients with the V30M TTR mutation from different tissues differ significantly in dimensions and overall appearance[58]. Amyloid fibres may also differ in the configurations of their lower-order components. Aβ and TTR-derived fibrils consist of interwound filaments, which occasionally split when deposited on surfaces[19, 44]. Whether the hierarchical nature of the structure is mirrored in a hierarchical assembly mechanism has yet to be discerned, mostly due to the fibril nucleus' obscurity. Fibrils are not static entities; Carula *et al.*[59] employed HDX combined with NMR and MS in order to probe peptide (PI3-SH3) stability in a fibrillar context. Experimental data led the authors to propose the existence of a dynamic equilibrium between fibrillar and soluble species giving rise to the observed behaviour. Overall, two processes may be distinguished during fibril growth: monomer/soluble oligomer addition and fibril breakage/re-association.

Deng *et al.*[60] have investigated the seeded fibril growth kinetics of modified TTR105-115 using FRET. FRET experiments have distinguished between two distinct processes occurring over the course of a day. EM analysis revealed that at the end of the first process peptides are partitioned between fibrils and flocculent aggregates but at the end of the second process (after ca. 10h of incubation) only bundled fibrils could be seen. Seeded growth thus proceeds by soluble peptide addition onto fibrillar seeds along with accumulation of larger aggregates, followed by incorporation of peptide in all other aggregation states into amyloid fibrils together with fibril maturation.

5.4.3 Fibril Maturation – Higher Order Structures

Further rearrangements may occur in aggregating solutions even after most soluble peptide has been incorporated into fibrils. Amyloid fibrils frequently appear forming larger aggregates and bundles *in vitro* (although some of these changes may be influenced by surface interactions and drying, processes necessary for microscopic

imaging). Amyloid plaques, large protein aggregates *per se*, are the most familiar encounters of disease-associated amyloid *in vivo*. Self-assembling higher-order structures may still contain appreciable order. Amyloid fibrils have been shown to form spherulites: spherical super-structures with semi-crystalline lamella radiating from their core, typically formed by synthetic polymers[61]. The formation of shperulites from amyloid-like fibrils has been detected in vivo[62, 63] and in vitro, by both biologically occurring proteins[62, 64] and small synthetic peptides[65].

Fibres with quasi-circular cross-section are not the only ordered structures that result from protein self-assembly; tape-like structures and broader ribbons may result by lateral association of fibrils[29, 65].

Fibrillation has an obvious effect on the bulk properties of aggregating mixtures. An increase in viscosity of concentrated peptide solutions to the point of forming a hydrated gel is most readily witnessed. Long-term fibril maturation has been shown for TTR105-115[31, 60]. Furthermore, it has been shown that pH can switch between different macroscopic phases in designed fibrillogenic peptide systems[65]. Fibrils, once grown, can self-associate further into higher order aggregates, a process which impacts micro- and macroscopic structure.

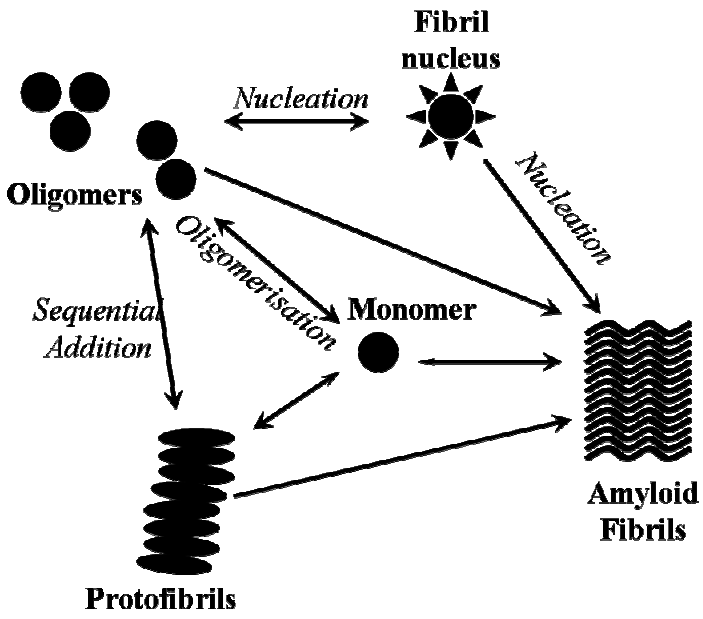


Figure 5.1 Scheme displaying different pathways to fibril formation. Monomers self-associate spontaneously into LMW oligomers which may either form intermediate higher-order aggregates (protofibrils) or/and undergo a conformational change into a structure that supports mature fibril growth, the latter possibly occurring by either or both incorporation of monomers and oligomeric aggregates. Alternatively, protofibrils may self-associate and re-arrange forming the first seed for mature fibrils. Fibril formation could proceed, in some cases, by downhill self-association. See text for a more detailed discussion including specific examples.

5.5 Insights into Fibrillogenesis by MS and IM-MS

The spontaneous formation of nm- and μm-sized structures from initially monomeric soluble peptide constituents proceeds in a complex manner. No single physical technique can adequately characterise aggregation products from each individual stage of fibrillogenesis; the peculiarities of amyloid formation thus necessitate the systematic coordination of fundamentally different approaches, making the field – as hopefully demonstrated above – intensely interdisciplinary. Observation of early monomeric and oligomeric species presents particular difficulties due to their dynamic nature (section 5.4). Since MS can readily detect species of different mass it is ideal for such a task. Identification of oligomeric aggregates by MS presents two main difficulties:

1. Higher-order oligomers may dissociate into monomer or lower-order oligomers
2. All peaks corresponding to $n\mathrm{M}^{nz}$ will appear as a single M^{z} peak in a mass spectrum.

Fortunately, these hurdles are not insurmountable. The ability to control the kinetic (and often internal) energy of ions in a mass spectrometer can be utilised to determine the extent at which (1) takes place; (2) can be resolved by IMMS as ions with the same m/z ratio but with different mobilities (mass, charge and conformation {1.23}) in an inert gas will be separated in a drift tube. Furthermore, oligomeric species – in contrast with higher-order aggregates – are amenable to all-atom simulations; thus ionic mobilities and the cross-sections they imply can serve as a simple yet powerful reference point between experimental data and simulated configurations. MS has attracted increasing attention as a potent technique in the field of amyloid formation, as a result of improvements in the transmission of non-covalently associated biomolecular complexes in a mass spectrometer (see for example Sobott *et al.*[66]). Some relevant work will now be reviewed in brief.

5.5.1 Amyloid-β Peptide

A large body of work, which has in part inspired the research on TTR105-115 reported herein, has been performed in the research group of Prof. M. T. Bowers in UCSB (in profuse collaboration with other researchers; see for example Teplow *et al.*[67]), utilising

IMMS to assess the configurations early species during amyloidogenesis of Aβ, α-synuclein[68] and IAPP[69]. Since their published work is focused on Aβ, we will focus on this peptide for demonstrating the capabilities of low-field IMMS on the early stages of amyloidogenesis. In 2005, Bernstein *et al.*[70] reported on the detection of early aggregate species of Aβ1-42 and the non-amyloidogenic Aβ1-42 F19P variant. The authors, electrospraying samples from 49.5% H_2O, 49.5% ACN and 1% NH_4OH, detected –2, –3, –4, –5 as well as $n2M^{5n}$ charge states for both peptides and an additional $n3M^{7n}$ charge state for Aβ1-42 F19P. ATDs resulting from these peaks were composite and the intensity of individual distributions varied with injection energy. Most notably, resolved species of the Aβ1-42 $n2M^{5n}$ ATD were assigned to dimer, tetramer, hexamer and dodecamer (n=1, 2, 3 and 6 respectively) yet the F19P mutant only contained dimer and tetramer. Furthermore, the absence of intermediate aggregates between putative hexamers and dodecamers led the authors to speculate that the hexamer was the "building block" of protofibril formation, a finding corroborated by cross-linking studies[52, 53, 71] among others. Additionally, removing higher-order aggregates by filtering immediately prior to MS analysis increased the aggregate population, indicative of higher-order aggregates incorporating smaller species. Baumketner *et al.*[72] subsequently performed extensive REM simulations on monomeric Aβ1-42 carrying a "physiological" –3 (–6+3) charge, both in vacuum and in GB solvent continuum. Representative geometries from GBS simulations were subsequently minimised *in vacuo* and grouped into three families, according to principle component analysis. Summation of four calculated steady-state ion transmission distributions (equation {1.30}), one for each "dehydrated" structural family and one for all solvent-free structures, resulted in an ATD model that could fit very well the one obtained experimentally. Although the authors admit that a number of alternative interpretations can be accommodated, should the one presented be correct, it implies that different structural families produced upon desolvation do not interconvert significantly during IMMS under the applied conditions, a phenomenon explained by "roughening" of the PES in low-dielectric environments. Structures of belonging to the three clusters of Aβ1-42 were very dissimilar (C_α-RMSD≈10Å) explaining the only partial agreement with NMR-derived solution

structures[67, 73] nevertheless two displayed some helical content, which was more accentuated in solvent-free simulations (where $\varepsilon_{diel} \equiv 1$), a finding which may explain the increased helical content of prefibrillar intermediates as determined by CD[56, 57]. IMMS combined with simulations can provide useful insights on the early events of fibrillogenic aggregation, including delineating monomer structure, oligomeric distributions and revealing mechanistic details during the aggregation process.

5.5.2 β_2 microglobulin

β_2 microglobulin (β2m) is another amyloidogenic β-sheet protein which has been studied extensively by mass spectrometry and ion mobility (FAIMS). Individuals undergoing renal dialysis contain significantly elevated amounts of β2m in their blood serum, which self-associates subsequently producing systemic dialysis-related amyloidosis. Like TTR, partial destabilisation of β2m's native fold is the first step of amyloidogenesis. *In vitro* this can be achieved by acid-mediated denaturation. Analysis of the monomeric protein's charge-state distribution over pH range has allowed Borysik *et al.*[74] to identify two partially folded states of β2m that become populated at pH<6 and quantify their relative populations as a function of pH. Only a native-like state, centred around the 7+ charge state, was observed at pH6; at pH4.8 a the first non-native population became apparent, centered around the 9+ and 10+ charge states; this population reached a maximum around pH3.6, where a second non-native population, centered around the 12+ charge state appeared, peaking at pH 2.4. This trend correlates well with pH-associated changes in fibril morphologies; at pH2.6 β2m forms long rigid amyloid-like fibrils whereas at pH3.6 it self-associates into much shorter, much more flexible worm-like fibrils (appearing similar to protofibrils of other proteins, described above) with both types displaying amyloid-like tinctorial properties[55]. FAIMS analysis of β2m ions at different pH values corroborated the conclusions based on deconvolution of charge-state distributions[75]. Whereas charge states associated with a particular ensemble (e.g. 7+, primarily associated with compact configurations) appeared as a single peak when travelling through the FAIMS device, charge states resulting from two conformational ensembles (such as the 8+ and 12+ charge states) separated into doublets, with pH-

dependent relative intensities[75]. Similar acid-unfolding effects were also observed for β2m carrying the V9A or F30A destabilising mutations, with the non-native states appearing at a pH higher than for wild-type β2m[74]. MS-based methods can also be used to distinguish the two observed pH-dependent fibril morphologies, as shown by Myers *et al.*[76]. Pepsin proteolysis of the two different fibril suspensions followed by MS analysis revealed that the two different fibrillar forms were characterised by different degrees of protection against cleavage giving rise to different peptide fragments. Proteolysis fragments were also found to be dependent on ionic strength. Notably, at pH 2.6, short incubations (15mins) with pepsin only resulted into one observable cleavage at V10, showing that most of the peptide is incorporated into the long and rigid amyloid fibrils. Conversely, worm-like fibrils were more susceptible to proteolysis. Not only can the two fibrillar morphologies be distinguished, but also details on the mechanism of self-assembly can be elucidated by MS approaches. Smith *et al.*[55] have monitored the time evolution of monomer concentration and oligomer distribution by MS in order to further understand β2m fibrillogenesis. At pH2.5 oligomers up to tetramer formed within 2 minutes of incubation; no higher-order oligomers were observed and after 12 hours of incubation the aggregates were depleted. On the other hand, at pH3.6 dimers are observed immediately after suspension, oligomers up to hexamer just after 2 minutes of incubation and further oligomers up to dekatriamer after one hour of incubation. After 8 and 12 hours of incubation, only monomers, dimers and trimers were observed, after which peptide was depleted at higher-order non-ionising assemblies. Monomer depletion (using bradykinin as an indicator of concentration) showed that self-assembly at pH3.6 follows a “downhill” aggregation mechanism, whereas at pH2.5 a more complex behaviour is evident. Variation ThT fluorescence during fibrillogenesis was in good agreement with the monomer depletion studies at pH3.6 and showed that at pH2.5 amyloidogenesis displays nucleation-and-growth kinetics. Thus it becomes evident that β2m fibrillogenesis proceeds via two competing pathways: partially unfolded monomers self-assemble spontaneously into aggregates with a worm-like morphology, yet formation of long, rigid amyloid fibres requires the formation of a high energy intermediate, associated with the critical nucleus and possibly linked with a different

conformational ensemble in solution than the flexible fibrils. These studies indisputably demonstrate the power of MS for characterising both steady-state and transient properties of amyloidogenic systems.

5.5.3 Transthyretin

Several MS studies have been performed on TTR, particularly assessing tetramer stability and the properties of non-native monomers which are thought, as outlined above, to be responsible for fibrillogenesis. Sobott *et al.*[66] have used TTR to demonstrate the capabilities of a modified QToF mass spectrometer to transmit large macromolecular assemblies; this was achieved by careful control of the pressures at the different sections of the instrument and adjustment of the quadrupole RF frequency. Such modifications allowed a more detailed description of the dissociation pathway of tetrameric TTR ions in an argon-filled collision cell to be carried out[77]. In the absence of organic co-solvents 15+ tetrameric TTR dissociated into a highly charged monomer (7+ to 10+) and a less charged trimer (5+ to 8+). This asymmetric dissociation pattern was also documented for the TTR octamer ions[77]; in fact, it is a feature frequently observed when performing CID on multimeric macromolecular polyionic complexes[78]. Multimeric TTR ions have also been analysed by TWIMS[*79]. Activating ions in-source by increasing the cone voltage induces significant broadening of arrival time distributions, indicating unfolding of the protein ions, with several unfolded species becoming apparent. Furthermore, TWIMS analysis can also reveal the presence of oligomeric species coinciding with native tetrameric charge states. Using much simpler instrumentation, Nettleton *et al.*[80] have followed the in-source dissociation of wild-type TTR and the V30M and L55P variants. The authors have found that the dissociation propensity of tetrameric TTR ions correlates with their amyloidogenicity and agrees well with previously published results on the tetrameric stability of these polypeptides. The stabilisation conferred by thyroxine binding was also witnessed by in-source CID[80], making MS an attractive technique for the screening of ligands that can stabilise a particular quaternary fold over all others, an application of great significance for therapeutics[41].

[*] Data reported by Ruotolo *et al.*, performed on a Synapt HDMS instrument.

5.6 Summary

Here MS studies on amyloid-forming systems, particularly those employing IMMS at one point or another, have been reviewed; yet the description does not by any means exhaust the volume of amyloid research that has used MS. For example, the oligomeric states of amyloidogenic insulin solutions has been followed by MS[81]. Additionally, other aspects of fibril formation and structure can be illuminated by MS. Knapman *et al.*[82] have used MS to determine the critical concentrations of synthetic fibrillogenic peptides. The structure of fully formed fibres can also be traced using MS-based methods, as shown already above for TTR. HDX has been employed as a probe of fibril structure for Aβ fibrils[51], which exhibit a protected core, and PI3-SH3 fibrils[59]; modelling of the exchange kinetics allowed significant insights on the dynamics of amyloid fibrils, with the emergence of a "molecular recycling" mechanism. Amyloidogenic polypeptides have now been analysed with all three ion mobility separators commonly interfaced with mass spectrometers. Evidently, MS and IMS are potent tools for identifying and characterising pivotal species associated with – often pathogenic – peptide self-assembly and can thus be employed for understanding the fundamental phenomena underlying self-association processes and assist the design of therapeutic agents for protein misfolding diseases or novel nanomaterials based on peptide as well as peptidomimetic building blocks.

5.7 Methods

Experimental IMMS Methods	
Solvent	49.5% H_2O 49.5% ACN 1% HCOOH
Concentration	834μM
Spray Voltage	2000
Cone Voltage range	45-95
Extraction Cone Voltage range	85-125
Injection Energy / V	31.5
Drift Voltages	60, 50, 40, 30, 25
Drift Pressure range / torr	3.070-3.200
Simulation Methods	
AMBER force field	ff99
MD time-step	1fs
Number of time-steps per run	5,000,000
Simulated Annealing Parameters	
MD time-step	1fs
No. of time-steps during high temperature MD	30,000
No. of time-steps during cooling	20,000
No. of cooling steps	20
High Temperature / K	800 (monomer), 600 (multimers)
Low Temperature / K	0
Total no. of iterations	500

5.8 Results & Discussion

5.8.1 Mass Spectrum of TTR Reveals the Presence of Oligomeric Aggregates

Aliquots of freshly prepared TTR105-115 stock solution* (8.34mM (10mgml^{-1}) in 89% H_2O, 10% ACN and 1% formic acid) incubated at room temperature for ca. 20-30 minutes were dissolved into 49.5% H_2O, 49.5%ACN and 1% formic acid to a final concentration of 834μM (1mgml^{-1})and analysed immediately on the MoQToF. Two independent measurements were performed, and typical data are displayed herein. Under these conditions, in ToF MS mode, TTR105-115 yields a mass spectrum like the one shown in figure 5.2a. The spectrum is dominated by $n\mathrm{M}^{n+}$ and $n\mathrm{M}^{2n+}$ peaks, accompanied by a wide range of oligomers. Transmission of the oligomeric species required an elevated pressure in the source region, a fact which has been reported previously[66]. Lower pressures yielded weak signal corresponding to monomers; furthermore similar samples sprayed on a QToF Ultima not possessing the source modifications of the MoQToF only generated monomer, dimer and trimer peaks.

In the absence of isotopic resolution of multiply-charged species there would be no indication about the oligomers that give rise to the spectrum. Fortuitously, one of the samples was contaminated with a potassium salt; the presence of potassium adduct series could thus confirm the charge assignment – based on predicted aggregate masses – and indicate the dominant oligomer populated in each peak in the spectrum (an example is shown in Figure 5.2b). The resulting assignment is shown in Figure 2.5a. All oligomers from dimer to dodecamer, with the exception of the endecamer, could be identified. Furthermore, the oligomer charge state distributions are of particular interest. Oligomer peaks carrying more charges than oligomers where never observed. This may be explained by the neutral character of TTR105-115, the N-terminus being the only primary amine. The endecapeptide possesses a net charge of 0; dilute (10-50μM) solutions in 49.5:49.5:1 H_2O:MeOH:HCOOH produced 1+ and 2+ ions. The fact that

* After days of incubation, concentrated solutions of TTR105-115 formed a gel, which could be compacted by centrifugation, indicating the presence of large amounts of insoluble fibril assemblies.

extended conformations tend to possess more charges in a mass spectrum is well documented[83]; thus the absence of highly charged assemblies indicates that the observed ions are characterised by compact geometries. The narrow charge-state distributions observed (with at most two charge states detected for each oligomer, although potential peak overlaps may have hindered a complete assignment) underline this point.

Do the observed oligomer distributions reflect their population in solution? Primarily it must be noted that a strictly linear correlation between concentration and m/z cannot be assumed. Even if an identical tendency to ionise is postulated, ions of different m/z are likely to be transmitted with different efficiency through the electrostatic lenses of a mass spectrometer. Yet aggregate formation can also occur during the electrospray process, as solvent evaporates from the droplets. Nevertheless, one would expect the population of oligomers resulting from such a 'non-specific' self-association process to decrease quasi-exponentially with oligomer size. However, this is not the case. Albeit high amounts of monomer being observed, the oligomer distribution implied by the mass spectrum of Figure 5.2a is centred on the tetramer, with both trimers and pentamers being populated. Such unconventional oligomer distributions have been observed previously by PICUP of amyloidogenic peptides[53, 54] and appear to be a trait of peptides which are able to self-associate. Such considerations increase our confidence in our data. Nevertheless, MS analysis may still *modify* the oligomeric distribution evidenced in the mass spectrum – not only due to in-source aggregation but also due to dissociation. Highly charged species may be rendered unstable in the absence of charge-screening solvent molecules, and dissociate into lower-order oligomers. Under the light of the experimental data it can be inferred that should such a process be taking place, it must be mediated by highly-charged *metastable* monomeric species. Indirect evidence of such a process is given by the CID-type fragments observed in the low molecular mass range of the spectra containing aggregates; b3-7, y7-3, x7, c4 and $a10^{2+}$ were detected (Figure 5.2d) – from the compact conformation of the monomer (see below) it can be deduced that an unusually high charge ($\geq$2) will considerably enhance fragmentation.

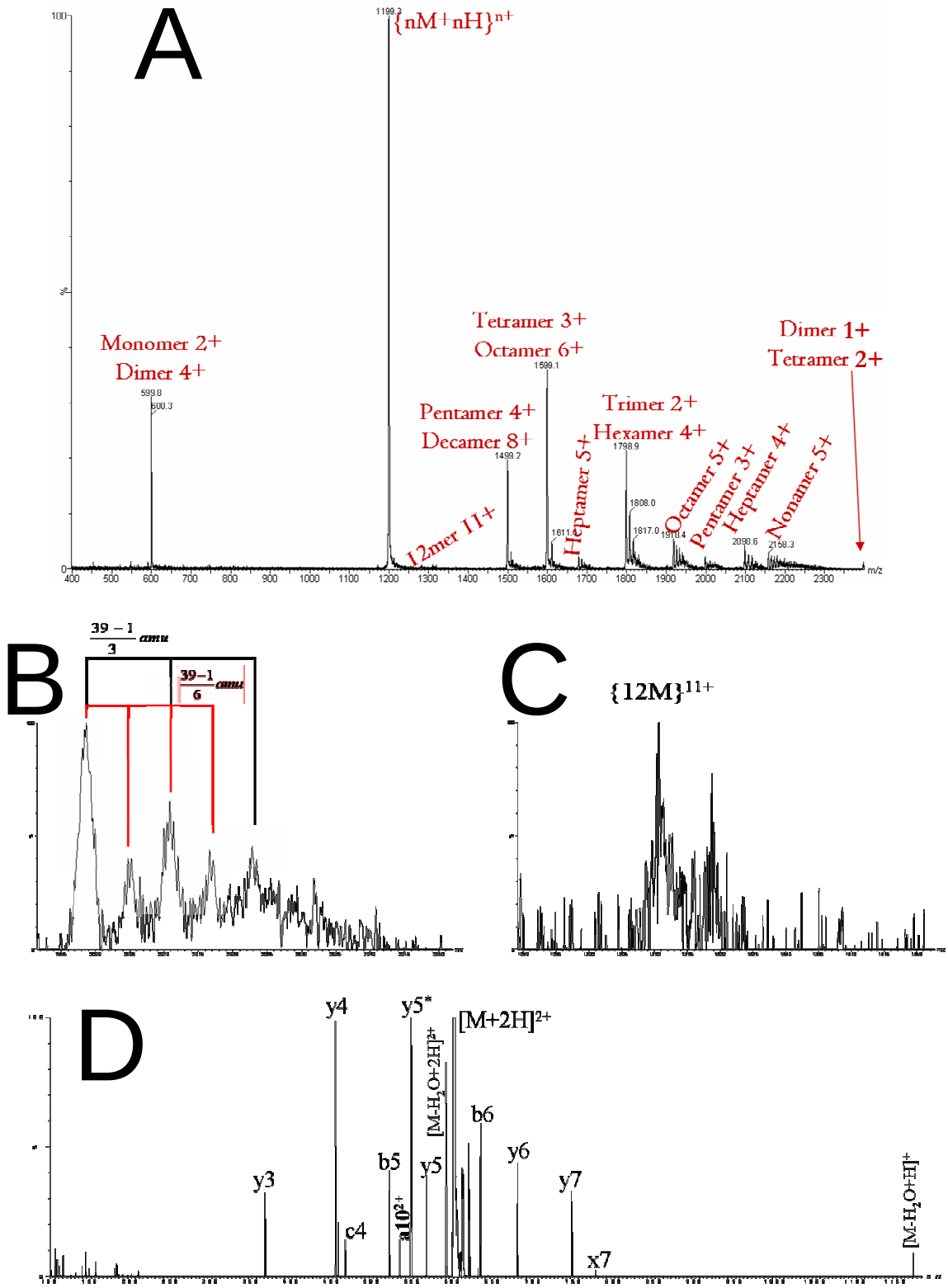


Figure 5.2 Mass spectra of aggregating TTR105-115 resuspended in 49.5%:49.5%:1% H_2O:ACN:HCOOH a) Full-range mass spectrum showing peaks arising from oligomeric ions. b) Zoom on the z = 3/5 peak displaying K^+ adducts – evidence for pentamers and decamers. c) Zoomed-in view of $[12M+11H]^{11+}$ peak. d) Low mass range of spectrum, showing most prominent CID-type fragments, presubeably formed by fragmentation of unstable highly charged monomers.

If detected oligomer ions are not produced by non-specific aggregation in the concentrated solutions, they must correspond to species present in the original aggregation mixture that survive dilution into a solution with 50% ACN or species which result from the disassembly of higher-order aggregates. In the first case, they could constitute prefibrillar oligomers whereas in the latter they could be conceived as building blocks of early-forming large aggregates. Interestingly, TTR105-115 protofilaments are thought to be composed from four stacked β sheets[84]; could the tetramer be a fibrillar constituent? As reviewed above, early aggregation stages of TTR105-115 is populated by species with different properties than mature fibrils[31, 60] allowing the observed oligomers to be read as prefibrillar states. To further assess the structure of the observed species we produced two classes of model structures, antiparallel β-sheets and globular aggregates and assessed their fitness by comparison to IMMS data.

5.8.2 Configurational Properties of Oligomeric Aggregates Probed by Ion Mobility

IMMS data were successfully collected for all oligomeric species apart from the dodecamer, for which signal was too weak. Oligomer cross sections are presented in Figure 5.3 and Table 5.2. As expected, a steady increase in collision cross section with oligomer order is observed, with the only exception being the 6+ decamer. Although the increase of cross section with oligomer order is monotonic, the rate of increase displays interesting features from which structural information can be inferred. Increases between monomer to tetramer and pentamer to nonamer are quasi-linear, yet changes between tetramer to pentamer and nonamer to decamer deviate significantly from linearity. Particular conformational rearrangements must take place to cause the pentamer, nonamer and decamer to be more compact than expected. Given that the most intense multimer peak is attributed to the 4+ tetramer ion, the data together are consistent with the theory seeing either tetrameric or pentameric states forming a stable conformation which may self-associate and thus become an important early intermediate during fibrillogenesis. However, it must be noted that – at least up to nonamer and decamer

formation, monomer addition cannot be excluded as a pathway for oligomer formation, since all intermediate oligomeric states are also populated – yet it is not ascertained either. The presence of intermediate oligomers between what can be believed to be the formation and self-association of a tetrameric or pentameric core aggregate, could also be attributed to aggregation during desolvation and gas-phase dissociation. Both mechanisms seem to be accommodated by the data and further research is necessary for distinguishing between the two.

Table 5.2

Momentum transfer cross-sections of TTR105-115 ions in He

Charge per monomer	Species	Collision cross-section Å^2
1	$[M+H]^+$	159.5
1	$[M+K]^+$	169.8
2	$[M+2H]^{2+}$	259.4
2	$[M+2Na]^{2+}$	265.3
1/2	$[2M+H]^+$	209.8
1	$[2M+H+Na]^{2+}$	289.9
2/3	$[3M+2H]^{2+}$	570.7
3/4	$[4M+3H]^{3+}$	630.3
3/4	$[4M+2H+Na]^{3+}$	480.3
3/5	$[5M+3H]^{3+}$	692.1
4/5	$[5M+4H]^{4+}$	682.4
4/5	$[5M+3H+Na]^{4+}$	716.7
4/5	$[5M+2H+2Na]^{4+}$	605.5
2/3	$[6M+4H]^{4+}$	800.8
4/7	$[7M+4H]^{4+}$	969.4
5/7	$[7M+5H]^{5+}$	943.8
5/8	$[8M+5H]^{5+}$	1091.9
5/9	$[9M+5H]^{5+}$	1122.2
3/5	$[10M+6H]^{6+}$	1016.2

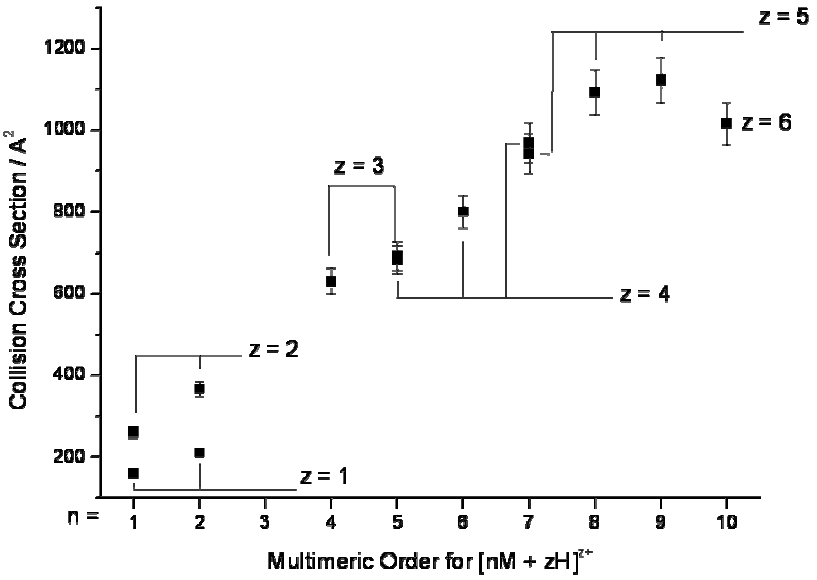


Figure 5.3 Experimental collision cross-sections of observed protonated TTR105-115 assemblies. Error bars were set to 5% of the cross-section estimate.

The 6+ decameric ions give rise to an unusually compact cross-section. The decamer cross-section was inferred from the apparent 3+ pentameric ions; if the ATDs of the two species overlap, the resultant cross-section estimate will be affected. The broad shape of the ATDs, pictured in Figures 5.4 and 5.5, imply the presence of multiple species, but, unfortunately, could not be clearly resolved. The ATDs themselves provide ample evidence for the presence of a mixture of oligomeric states in many cases. Whereas the 2+ peak is unimodal, the 1+ ion displays a late-arriving tail after IM separation. This feature – occasionally encountered with Trp cage 2+ ions (Chapter 3) – indicates monomer dissociation from higher-order aggregates while they are drifting. Thus, clear separation of monomers and oligomers of the $n\mathrm{M}^{n+}$ kind is hurdled. z = 1/2, 4/3, 5/4, 5/3 and 8/5 ions all separate into more than one population, which can most aptly be assigned as aggregation states, whereas z = 7/4 and 9/5 ions appear as single, but broadened distributions. The tetramer case is particularly interesting. z = 4/3 ions drifting at 40V separate into two broad peaks: one late-arriving population in which three features are discernible, and one early-arriving distribution. The assignment which

we propose is as follows: 3+ tetramers and 6+ octamers are grouped in the late-arriving peak, whereas 9+ dodecamers and possibly 12+ decahexamers, populate the early peak. Classifying octamers into the first peak follows from the fact that 5+ octamers are also characterised by longer rather than shorter drift times, arriving between 1.9 and 2.9μs at 40V and 3torr He, whereas the putative dodecamer arrives around 0.8μs under the same conditions. z = 5/4 and 5/3 ATDs compare qualitatively very well, being composed of a main peak flanked by several minor species, the most populated of which being at shorter times. These early-arriving species can be attributed to decamers, whose presence is readily attested by their potassium adducts in the mass spectrum (Figure 5.2). Charge and adductation influence the cross sections of different oligomers in different ways. Whereas 1+ monomer and dimer ions are characterised by a significantly smaller cross section than the corresponding 2+ species, addition of a further charge on pentamers or heptamers has a minimal effect, illustrating that the *average charge per monomer* exerts a more potent effect on conformation than absolute charge *per se*. Oligomeric species form potassium adducts quite readily, underlining the neutral character of the peptide. Yet contrary to expectations, addition of a single Na^+ ion tends to cause a small *increase* in cross-section, apart of the tetramer which becomes compact. Similarly, pentamer:Na adduct is more extended than the isolated pentamer; yet addition of a second sodium ion brings about a significant structural change. Based on this finding, the hypothesis that addition of metal cations leads to eventual destabilisation of a particular structure formed by self-associated TTR105-115 peptides can be formulated and investigated further by additional experiments in the future. Interestingly, it has been long known that ionic strength influences the kinetics of fibril formation for different systems (see e.g. Harper *et al.*[43]).

IMMS alone can provide with relative rapidity a plethora of information on these dynamic self-assembling systems. We next use molecular modelling to assess the conformations adopted by the observed oligomers.

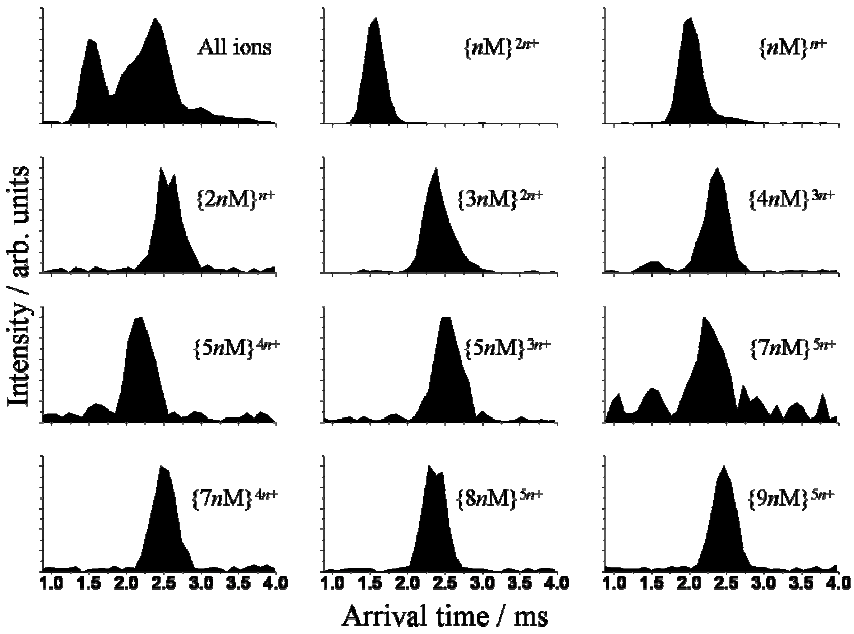


Figure 5.4 Arrival time distributions of main observed protonated ions. Labels are in $\{mn\mathrm{M}\}^{zn+}$ *m* being the smallest oligomer order carrying an integer charge z giving rise to the observed ion; n indicates the fact that higher-order oligomers may populate the ATD, as readily attested for most species, giving rise to broadened complex peak shapes.

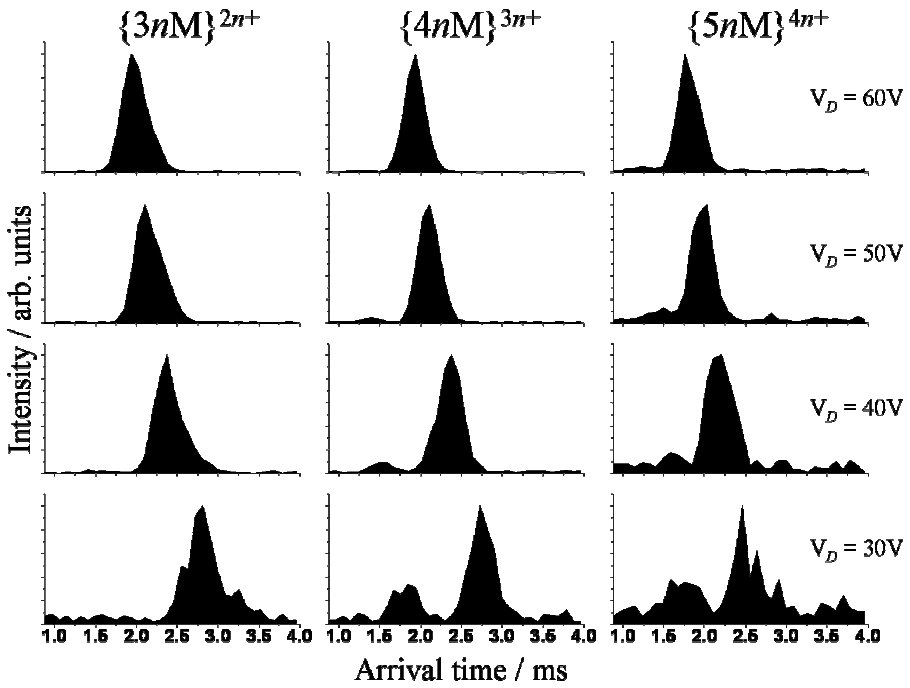


Figure 5.5 Arrival time distributions of (from left to right) z = 2/3, 3/4 and 4/5 peaks, at drift voltages of 60, 50, 40 and 30V, from top to bottom. Species that are unresolvable at higher drift voltages become apparent at lower values, with simultaneous loss in intensity due to diffusion. Ions arriving at earlier times experience less depletion due to drift.

5.9.1 Modelling Strategy

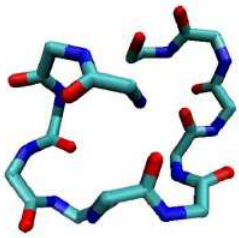
What are the structural characteristics of the detected oligomers? Do they represent stable oligomeric species in solution? Or are they offspring of larger, more labile aggregates? Although exact conformation cannot be definitely determined from experimental data alone, as outlined in previous chapters, IM-MS is particularly powerful in screening between different candidate structural families[67]. If the observed oligomers are dissociates of fibrillar matter, the latter containing characteristic β-sheet structure, they may well preserve this β character *in vacuo*. If the observed ions represent genuine stable solution-phase species in would still be interesting to examine the possibility that they adopt a β-sheet conformation. Since β-strands are completely extended, collapsed conformations represent an altogether different structural family. B-sheet structures where based on a protofilament model based on the solid-state NMR structure of TTR105-115 in fibrillar form[26, 27] arranged in an antiparallel manner. Species from monomer up to hexamer were modelled. Given the size of the larger oligomers, we restricted our research to $n\mathrm{M}^{(n-1)+}$ charge states since they would tend to present less overlap with lower-order oligomers in a mass spectrum[*]. Thus, for each oligomer consisting of *n* monomers, *n*-1 peptides will contain a charge of +1 and one will be neutral. Charged peptides were protonated at the N-terminus, the only free amine group, and contained a neutral C-terminus; in neutral peptides both termini were charged, yielding a net charge of 0. In the initial structure of oligomers larger than dimer the central strand was chosen to be zwitterionic. When an internal strand is zwitterionic, the anionic C-terminus will be flanked by positively charged N-termini of neighbouring strands, augmenting the stability of the arrangement. Initial structures (model β-sheets procured by Dr. Cait MacPhee) were energy-minimised according to the protocol outlined in section 2.2.2.3, were heated gently to 350K and subjected to 5ns of MD. The same initial structures were processed by re-iterations of the simulated annealing algorithm to sample compact conformations. Using identical force-field parameters for both families allows direct comparison of the energies, simplifying the interpretation of

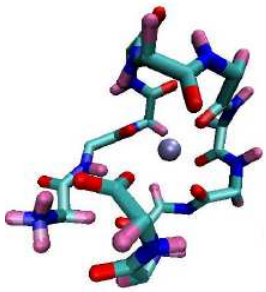
[*] Although calculations were performed before experimental data were obtained, it was proven that dimers, pentamers, tetramers and pentamers all appear as $n\mathrm{M}^{(n-1)+}$ ions in a mass spectrum, showing that our prediction was quite apt in these cases.

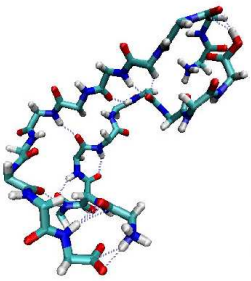
the relative stability of different structures. Simulation parameters are presented in Table 5.2. Numerically estimated cross-sections, calculated with mobcal, were compared with the experimental data, whereas further analysis of the simulations led to interesting insights on the properties of aggregate configurations in a minimal, solvent-free environment.

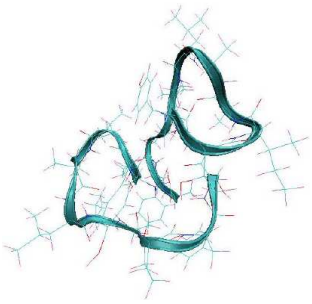
5.9.2 Monomer Structure and Effect of Cation Binding

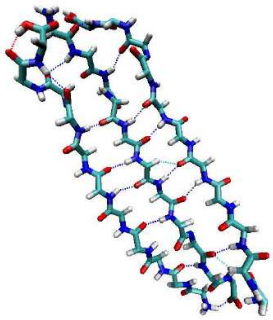
As often seen for several oligopeptides, the gas-phase configuration of isolated TTR105-115 monomers appears to be driven entirely by charge solvation. With the proton located at the N-terminus, 1+ TTR105-115 ions adopt a compact conformation with the protonated group located at the centre of the structure, interacting with backone carbonyl oxygen atoms (Figure 5.6). During 5ns MD at 300K the monomer preserves its basic topological features. Notably, no appreciable helical structure is evident neither among low-energy simulated annealing-generated structures nor during MD. Simulated annealing of the $[M+Na]^+$ adduct produces even more compact conformations (only the monomeric peptide:sodium complex was simulated). The sodium ion in this case forms a structural core, being extensively chelated by carbonyl oxygen atoms (Figure 5.6). The effect of sodium binding is reflected in the theoretical collision cross-sections, calculated with the trajectory method[85], which are 270Å^2 for the free peptide and 254Å^2 for the peptide:sodium complex. Contrary to this prediction, experimentally observed Na^+ adducts are slightly more extended, implying the presence of particular structural characteristics which binding of a single cation cannot overturn.

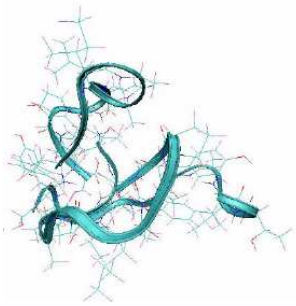
A 

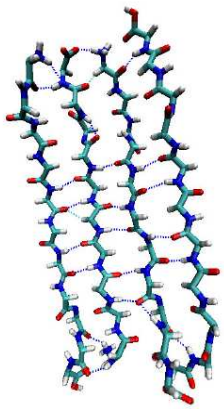
 B

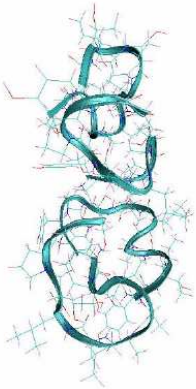
C 

 D

E 

 F

G 

 H

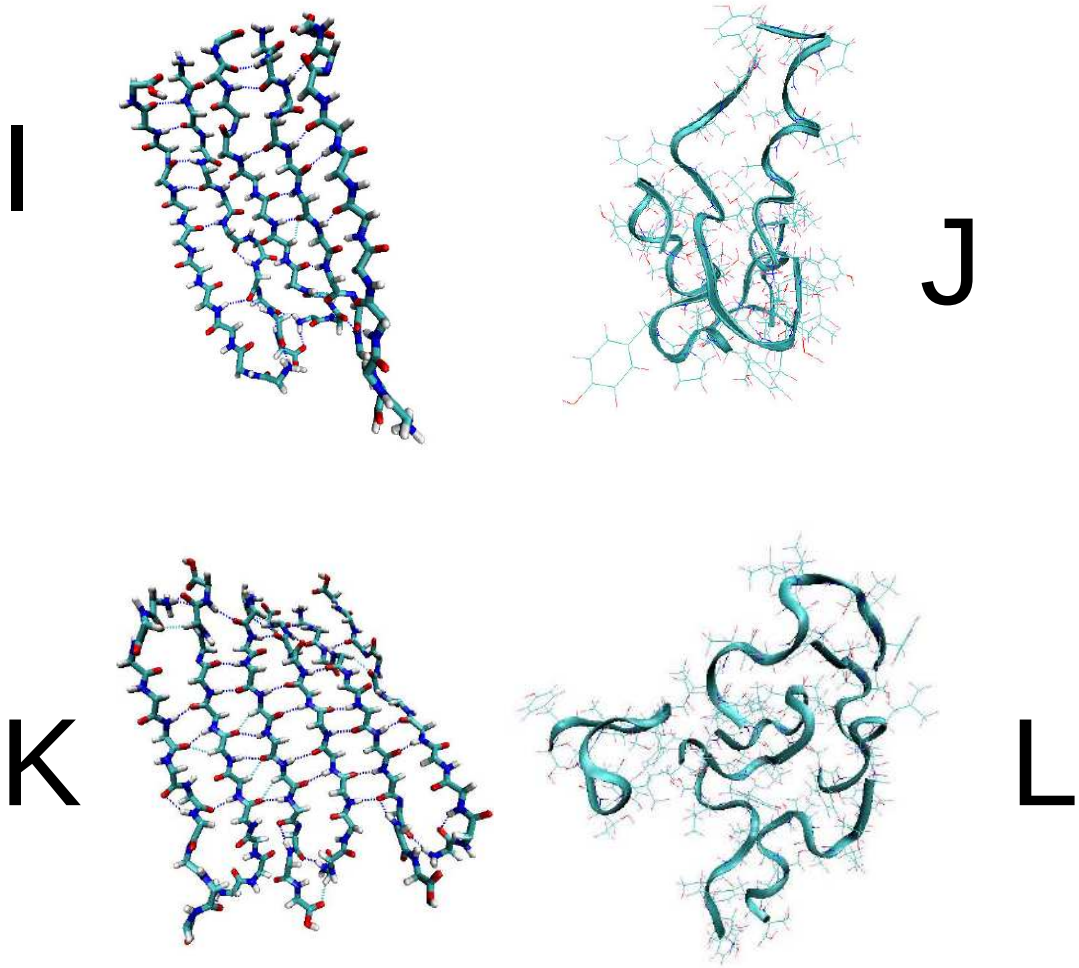


Figure 5.6 Simulated geometries of TTR105-115 monomer and oligomers; structures resulting at the end of 5ns MD at 300K starting from an antiparallel β-sheet arrangement followed by a low-energy structure sampled by the simulated annealing algorithm. a) monomer (sim. annealing); b) sodiated monomer (sim. annealing); c-d) dimer e-f) trimer g-h) tetramer i-j) pentamer k-l) hexamer.
Evidently, peptide geometries generated by simulated annealing are more collapsed yet individual peptides gradually become more extended in larger assemblies. Antiparallel β-sheet configuration is stable during MD of trimers to hexamers – dimers form a more compact 'β-bulge' with many backbone H-bonds remaining intact.

5.9.3 Stability of Antiparallel β-sheets in vacuo

After assigning force-field parameters to the different systems using the *xleap* program included in the AMBER package, the stability of the β-sheet conformation was monitored by running 5ns MD for each oligomer. MD trajectories were inspected visually, representative structures being shown in Figure 5.6. The calculated collision cross-section is used as an indicator of the global configuration of the peptides in addition to the radius of gyration and N-to-C-terminal distances, order parameters which can differentiate between the conformations of individual peptides in an oligomer. As can be easily deduced from Figures 5.6 and 5.7, dimers do not preserve their initial β-sheet conformation. Nevertheless, many of the backbone hydrogen bonds remain intact, the resulting conformation being a β-bulge, which remains stable for the majority of the simulation time. in the course of extensive computational work on dimeric TTR105-115 in implicit solvent, Li *et al.*[86] have found both parallel and antiparallel β-sheets as well β-bulges to feature among the low energy conformations of the dimer, the parallel orientation gaining precedence over the antiparallel orientation. Due to the absence of solvent in our simulations long-range electrostatics become much more significant, rendering the β-bulge conformation more stable than β-sheet structures. Conversely all higher-order oligomers retained their β-sheet character during the MD simulation. Interactions between neighbouring termini (neutral or charged), caused some deformation of the structures near the termini; such effect is quite pronounced in the case of the trimer (Figure 5.6).

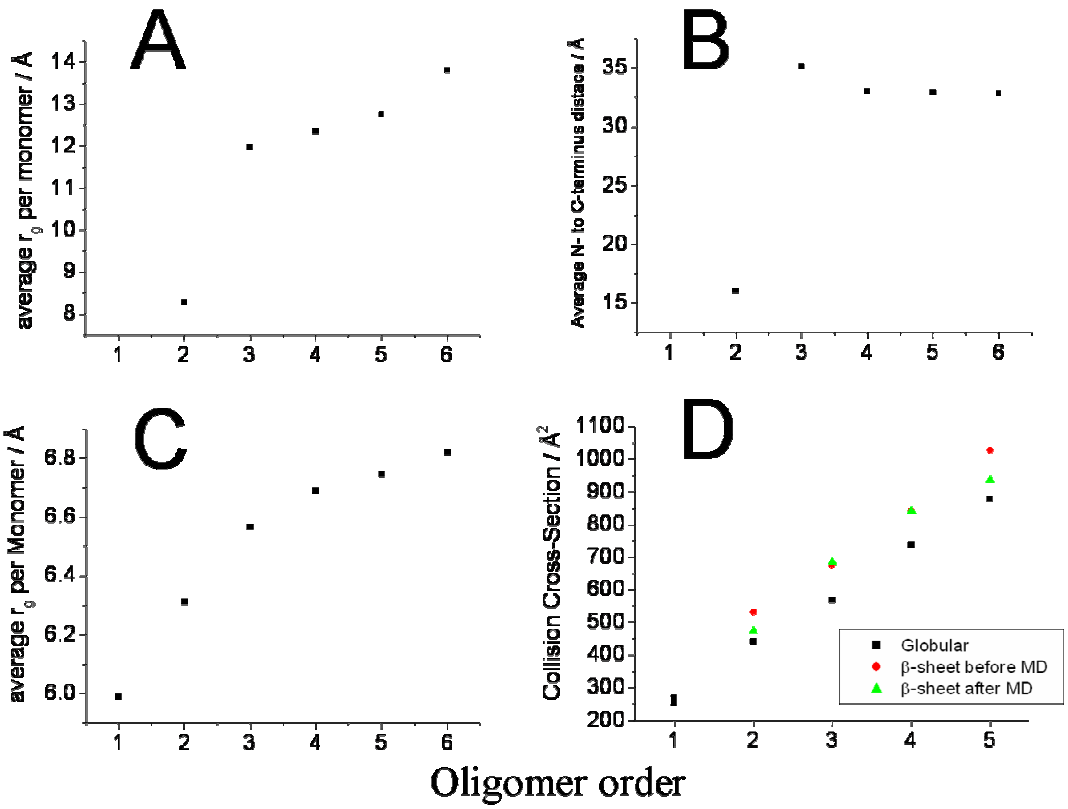


Figure 5.7 a) Average radius of gyration (r_g) per monomer during MD of simulated oligomers in antiparallel β-sheet configuration. b) Average intramolecular distance between peptide termini during MD of simulated oligomers. c) Average r_g of globular oligomers generated by simulated annealing. d) Comparison of collision cross sections calculated for globular and β-sheet oligomers before and after relaxation MD. Averages from MD simulations ignored the first 2.5ns of the simulation.

5.9.4 Amorphous Globular Aggregates Probed by Simulated Annealing

The simulated annealing schedule described in section 2.2.2.4 was iterated 500 times for each multimer – averages were rescaled according to equation {1.19}. Even low-energy simulated annealing geometries did not contain extensive β-sheet character and so served as models for collapsed, amorphous structures. Lowest-energy configurations are shown in Figure 5.6. The radius of gyration of *each individual monomer* was calculated, the average radius of gyration per monomer being presented in Figure 5.7a. It can be clearly seen that monomers become more extended as the aggregate size increases; this is indicative of an increase of the extent of intermolecular interactions (relative to

intramolecular contacts) for larger aggregates. Furthermore, an abrupt change in the rate increase is observed between trimer and tetramer. An average increase in radius of gyration of 0.29Å per monomer is observed between monomer and trimer which drops to 0.07Å per monomer between tetramer and hexamer. It appears then that significant internal rearrangements take place as smaller oligomers are formed in order to accommodate more intermolecular interactions, but once a tetramer is formed this process is nearly complete. Further rearrangements will be required to reach the most favourable structure – as only short annealing times were allowed and all parameters were kept constant for all oligomers, we do not expect the simulation to have converged.

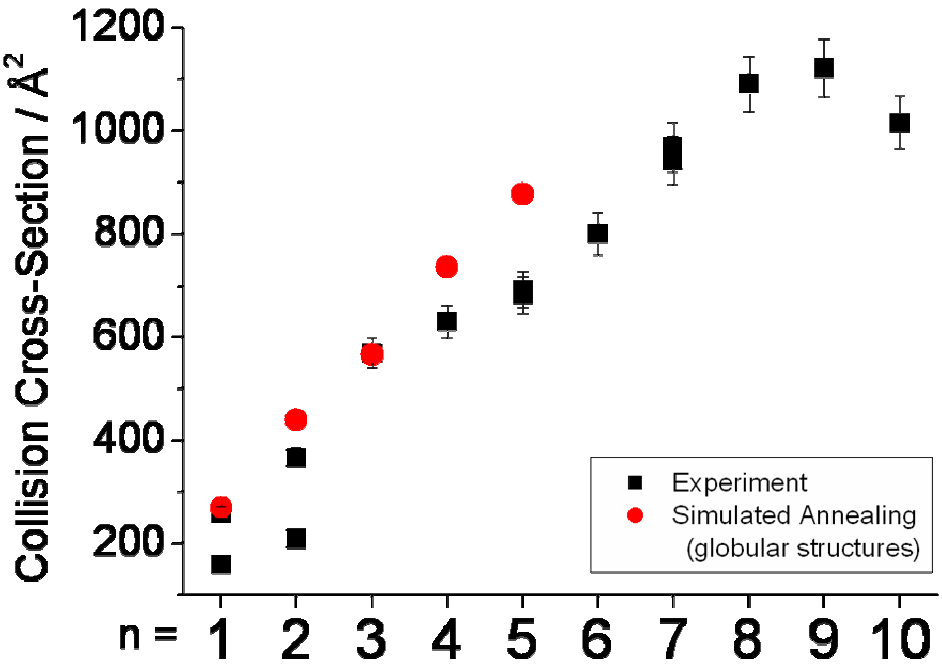


Figure 5.8 Comparison between experimentally and computationally derived collision cross-sections. Experimental momentum transfer cross-sections are consistently smaller than corresponding to globular structures, showing that experimentally observed species adopt compact conformations.

5.10 Observed oligomers Adopt a Compact Globular Conformation

Calculated collision cross-sections for both classes of conformations, shown in Figure 5.7, display two noteworthy qualitative trends:

- collision cross-sections of simulated multimers increase in a linear fashion (with a rate of 165Å^2 and 148Å^2 per monomer added for β-sheet and collapsed structures respectively)
- the difference in the rate of increase is greater than 10% than the rate of the β-sheet conformation (over the range of oligomers studied the two conformational classes differ in cross-section on average by 113Å^2)

showing that the two structures are distinguishable by IMMS. The rate of increase in cross-section with multimer order for the experimentally observed species does not exceed the one predicted computationally by simulated annealing.

By comparing experimental and calculated momentum transfer cross-sections (Figure 5.8) it can be seen that detected oligomer ions have lower cross-sections than predicted for structures generated by simulated annealing. Moreover, experimental cross-sections increase with oligomeric order at a rate lower than that predicted for both classes of structures. Generally, one can safely conclude that oligomers adopt compact conformations at least in a solvent-free milieu. Due to significant quantitative differences between experimental and calculated collision cross-sections structural assignments of observed species to simulated structures cannot be performed. Several factors may give rise to such a discrepancy. From an experimental point of view, and already discussed in Chapters 2 and 3, injection energy and the presence of unresolved higher-order oligomeric ions under the ATDs analysed contribute to reducing the cross-section estimate – the decrease being more pronounced for highly-charged species. On the other hand, there may be a systematic error in the way the collapsed oligomers were generated. Indeed, keeping parameters unaltered does impede the convergence of structures to the global minimum (as shown in section 2.2.2.4). Furthermore, failure of the simulations to reproduce the smaller relative cross-section increase between tetramer and hexamer points towards the fact that conformations sampled can only serve as a

low-resolution guide to peptide arrangement. Thus for a more exact treatment of systems of such complexity more sophisticated computational methods than the ones employed herein are necessary. Yet even after paying due respect to these limitations, the difference between collapsed and β-stranded structures is clear and we conclude that observed oligomers belong to the former class.

5.11 Implications for TTR105-115 Fibrillogenesis

Several different pathways (illustrated in Figure 5.1 for amyloidogenic peptides) result in self-assembly of peptide systems[67, 87]. If each consecutive association step is energetically favourable, basic units will aggregate spontaneously without an appreciable lag phase, at a rate determined by diffusion rates and concentration of monomeric and multimeric species. However if high-energy intermediates separate monomeric/oligomeric from large assemblies, aggregation will proceed with nucleation-and-growth kinetics. Above a certain critical concentration, the, often hydrophobic, amyloidogenic peptides begin self-associating into oligomeric species. Fibrillar matter may form by both competing mechanisms, depending on environmental conditions and peptide sequence, as reported for β2-microglobulin[55]. Moreover, system-specific structural characteristics may render certain oligomeric species more stable than others, leading to such states being preferably occupied early stages of fibrillogenesis; analogous behaviour gives rise to ‘magic numbers’ of atomic and molecular clusters. Furthermore, a particular oligomeric species, instead of the monomer, may be the building unit of one class of aggregates – as suggested for Aβ protofibrils[67].

Observation of oligomeric ions up to dodecamer by MS from suspensions of aggregating solutions of 8.34mM TTR105-115 clearly demonstrate the ability of the peptide to spontaneously self-assemble, establishing an equilibrium between oligomeric states during the initial stages of fibrils formation. Oligomer populations, as inferred by the mass spectrum, are irregular and display a preference for the formation of tetramers over other oligomers. IMMS separation of z = ¾ ion revealed the presence of several oligomeric species. Notably, both sets of simulations attribute special properties to the tetramer, which gives rise to the most populated oligomeric ion and can self-associate

into aggregates of 4*m* order (where *m* = 1, 2, 3, …). Tetramers could form building blocks of larger aggregates which may be on- or off-pathway to fibril formation. Alternatively, tetramers could represent the largest oligomeric state before which aggregates form rapidly. However the presence of pentamer ions at significant intensity contradicts any hypothesis based on their transient state. Additionally, the observation of a wide range of oligomers does not rule out downhill aggregation taking place by monomer addition, even if the presence of 'intermediate' aggregates can be interpreted as being due to gas phase effects. Intriguingly, the dimensions of TTR105-115 potofilaments are consistent with them being composed of four stacked β-sheets – could it be that the tetrameric core undergoes further rearrangements to support fibril formation by simultaneous growth of those four β-sheets?

Deduction of collision cross-sections from drift times of oligomer ions in the ion mobility cell and comparison with theoretical estimates based on antiparallel β-sheet and globular structures showed that in fact observed oligomeric aggregates are globular and remarkably compact. Evidence for compact conformations can also be inferred from the narrow charge-state distributions, yet contraction due to removal of solvent and self-solvation cannot be excluded for causing such a phenomenon. Analysis of structures generated by simulated annealing suggests that isolated ionised droplets of solvent-free peptides, TTR105-115 monomers gradually re-arrange, replacing intramolecular with intermolecular non-covalent interactions. We can speculate that more thorough conformational sampling will reveal structures with greater regularity and better correspondence with experimental cross-sections.

TTR105-115 has the potential of being an important system for amyloid research, either by being used as a convenient model system for studying amyloid fibril formation or by being further modified to engineer new functionalities. IMMS can be very profitable in the quest of investigating the conformations of oligomeric species that become populated during the initial stages of aggregation, before they are incorporated into larger aggregates. The kinetics of self-association – particularly the transitions pertaining to early oligomerisation states – remain to be elucidated; with the aid of external quantitation temporal fluctuations in early aggregating species has been achieved by

MS[55]; the temporal changes in oligomeric distributions and structures of observed aggregates of TTR105-115 define the direction for future investigations on this system. Moreover, modifications on peptide structure, aggregation conditions and addition of ligands that either accelerate or impede fibril formation can be monitored by mass spectrometry via the impact these factors have on oligomeric distributions; implementation of IMMS is capable of revealing conformational changes associated with these events, leading to further advances on the field of protein folding and self-assembly, its clinical relevance and technological prospects.

5.12 References

1. Puchtler, H.; Sweat, F., A REVIEW OF EARLY CONCEPTS OF AMYLOID IN CONTEXT WITH CONTEMPORARY CHEMICAL LITERATURE FROM 1839 TO 1859. *Journal of Histochemistry & Cytochemistry* **1966,** *14* (2), 123-&.
2. Sipe, J. D., AMYLOIDOSIS. *Annual Review of Biochemistry* **1992,** *61*, 947-975.
3. Chiti, F.; Dobson, C. M., Protein misfolding, functional amyloid, and human disease. *Annual Review of Biochemistry* **2006,** *75*, 333-366.
4. Fowler, D. M.; Koulov, A. V.; Balch, W. E.; Kelly, J. W., Functional amyloid - from bacteria to humans. *Trends in Biochemical Sciences* **2007,** *32* (5), 217-224; Kelly, J. W.; Balch, W. E., Amyloid as a natural product. *Journal of Cell Biology* **2003,** *161* (3), 461-462.
5. Serio, T. R.; Cashikar, A. G.; Kowal, A. S.; Sawicki, G. J.; Moslehi, J. J.; Serpell, L.; Arnsdorf, M. F.; Lindquist, S. L., Nucleated conformational conversion and the replication of conformational information by a prion determinant. *Science* **2000,** *289* (5483), 1317-1321.
6. Fowler, D. M.; Koulov, A. V.; Alory-Jost, C.; Marks, M. S.; Balch, W. E.; Kelly, J. W., Functional amyloid formation within mammalian tissue. *Plos Biology* **2006,** *4* (1), 100-107.
7. Kenney, J. M.; Knight, D.; Wise, M. J.; Vollrath, F., Amyloidogenic nature of spider silk. *European Journal of Biochemistry* **2002,** *269* (16), 4159-4163; Slotta, U.; Hess, S.; Spiess, K.; Stromer, T.; Serpell, L.; Scheibel, T., Spider silk and amyloid fibrils: A structural comparison. *Macromolecular Bioscience* **2007,** *7* (2), 183-188.
8. Shao, Z. Z.; Vollrath, F., Materials: Surprising strength of silkworm silk. *Nature* **2002,** *418* (6899), 741-741.
9. Smith, J. F.; Knowles, T. P. J.; Dobson, C. M.; MacPhee, C. E.; Welland, M. E., Characterization of the nanoscale properties of individual amyloid fibrils. *Proceedings of the National Academy of Sciences of the United States of America* **2006,** *103* (43), 15806-15811.
10. Channon, K.; MacPhee, C. E., Possibilities for 'smart' materials exploiting the self-assembly of polypeptides into fibrils. *Soft Matter* **2008,** *4* (4), 647-652.
11. Reches, M.; Gazit, E., Casting metal nanowires within discrete self-assembled peptide nanotubes. *Science* **2003,** *300* (5619), 625-627.
12. Pastor, M. T.; Esteras-Chopo, A.; de la Paz, M. L., Design of model systems for amyloid formation: lessons for prediction and inhibition. *Current Opinion in Structural Biology* **2005,** *15* (1), 57-63.
13. Hamilton, J. A.; Benson, M. D., Transthyretin: a review from a structural perspective. *Cellular and Molecular Life Sciences* **2001,** *58* (10), 1491-1521.
14. Westermark, P.; Sletten, K.; Johansson, B.; Cornwell, G. G., FIBRIL IN SENILE SYSTEMIC AMYLOIDOSIS IS DERIVED FROM NORMAL TRANSTHYRETIN. *Proceedings of the National Academy of Sciences of the United States of America* **1990,** *87* (7), 2843-2845.

15. Hornberg, A.; Eneqvist, T.; Olofsson, A.; Lundgren, E.; Sauer-Eriksson, A. E., A comparative analysis of 23 structures of the amyloidogenic protein transthyretin. *Journal of Molecular Biology* **2000,** *302* (3), 649-669.
16. Cornwell, G. G.; Sletten, K.; Johansson, B.; Westermark, P., EVIDENCE THAT THE AMYLOID FIBRIL PROTEIN IN SENILE SYSTEMIC AMYLOIDOSIS IS DERIVED FROM NORMAL PREALBUMIN. *Biochemical and Biophysical Research Communications* **1988,** *154* (2), 648-653.
17. Lai, Z. H.; Colon, W.; Kelly, J. W., The acid-mediated denaturation pathway of transthyretin yields a conformational intermediate that can self-assemble into amyloid. *Biochemistry* **1996,** *35* (20), 6470-6482; Lashuel, H. A.; Lai, Z. H.; Kelly, J. W., Characterization of the transthyretin acid denaturation pathways by analytical ultracentrifugation: Implications for wild-type, V30M, and L55P amyloid fibril formation. *Biochemistry* **1998,** *37* (51), 17851-17864.
18. Pettersson, T.; Carlstrom, A.; Jornvall, H., DIFFERENT TYPES OF MICROHETEROGENEITY OF HUMAN THYROXINE-BINDING PREALBUMIN. *Biochemistry* **1987,** *26* (14), 4572-4583.
19. Cardoso, I.; Goldsbury, C. S.; Muller, S. A.; Olivieri, V.; Wirtz, S.; Damas, A. M.; Aebi, U.; Saraiva, M. J., Transthyretin fibrillogenesis entails the assembly of monomers: A molecular model for in vitro assembled transthyretin amyloid-like fibrils. *Journal of Molecular Biology* **2002,** *317* (5), 683-695.
20. Quintas, A.; Saraiva, M. J. M.; Brito, R. M. M., The tetrameric protein transthyretin dissociates to a non-native monomer in solution - A novel model for amyloidogenesis. *Journal of Biological Chemistry* **1999,** *274* (46), 32943-32949; Quintas, A.; Vaz, D. C.; Saraiva, M. J.; Brito, R. M. M., Partial unfolding of the non-native monomer is the key event in transthyretin amyloidogenesis. *Amyloid-Journal of Protein Folding Disorders* **2001,** *8*, 106-106; Quintas, A.; Vaz, D. C.; Cardoso, I.; Saraiva, M. J. M.; Brito, R. M. M., Tetramer dissociation and monomer partial unfolding precedes protofibril formation in amyloidogenic transthyretin variants. *Journal of Biological Chemistry* **2001,** *276* (29), 27207-27213.
21. Gustavsson, A.; Engstrom, U.; Westermark, P. In *FORMATION OF FIBRILS BY NORMAL TRANSTHYRETIN AND SYNTHETIC TRANSTHYRETIN FRAGMENTS INVITRO*, 6th International Symp on Amyloidosis, Oslo, Norway, Aug 05-08; Natvig, J. B.; Forre, O.; Husby, G.; Husebekk, A.; Skogen, B.; Sletten, K.; Westermark, P., Eds. Oslo, Norway, 1990; pp 591-594.
22. Gustavsson, A.; Engstrom, U.; Westermark, P., NORMAL TRANSTHYRETIN AND SYNTHETIC TRANSTHYRETIN FRAGMENTS FORM AMYLOID-LIKE FIBRILS INVITRO. *Biochemical and Biophysical Research Communications* **1991,** *175* (3), 1159-1164.
23. Jarvis, J. A.; Craik, D. J.; Wilce, M. C. J., X-RAY-DIFFRACTION STUDIES OF FIBRILS FORMED FROM PEPTIDE-FRAGMENTS OF TRANSTHYRETIN. *Biochemical and Biophysical Research Communications* **1993,** *192* (3), 991-998.
24. Serpell, L. C.; Sunde, M.; Blake, C. C. F., The molecular basis of amyloidosis. *Cellular and Molecular Life Sciences* **1997,** *53* (11-12), 871-887.
25. Jarvis, J. A.; Kirkpatrick, A.; Craik, D. J., H-1-NMR ANALYSIS OF FIBRIL-FORMING PEPTIDE-FRAGMENTS OF TRANSTHYRETIN. *International Journal of Peptide and Protein Research* **1994,** *44* (4), 388-398; Jarvis, J. A.; Craik, D. J., C-13 NMR RELAXATION STUDIES OF MOLECULAR-MOTION IN PEPTIDE-FRAGMENTS FROM HUMAN TRANSTHYRETIN. *Journal of Magnetic Resonance Series B* **1995,** *107* (2), 95-106.
26. Jaroniec, C. P.; MacPhee, C. E.; Astrof, N. S.; Dobson, C. M.; Griffin, R. G., Molecular conformation of a peptide fragment of transthyretin in an amyloid fibril. *Proceedings of the National Academy of Sciences of the United States of America* **2002,** *99* (26), 16748-16753.
27. Jaroniec, C. P.; MacPhee, C. E.; Bajaj, V. S.; McMahon, M. T.; Dobson, C. M.; Griffin, R. G., High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America* **2004,** *101* (3), 711-716.
28. MacPhee, C. E.; Dobson, C. M., Chemical dissection and reassembly of amyloid fibrils formed by a peptide fragment of transthyretin. *Journal of Molecular Biology* **2000,** *297* (5), 1203-1215.
29. Jimenez, J. L.; Nettleton, E. J.; Bouchard, M.; Robinson, C. V.; Dobson, C. M.; Saibil, H. R., The protofilament structure of insulin amyloid fibrils. *Proceedings of the National Academy of Sciences of the United States of America* **2002,** *99* (14), 9196-9201.

30. Mesquida, P.; Riener, C. K.; MacPhee, C. E.; McKendry, R. A., Morphology and mechanical stability of amyloid-like peptide fibrils. *Journal of Materials Science-Materials in Medicine* **2007,** *18* (7), 1325-1331.
31. Dirix, C.; Meersman, F.; MacPhee, C. E.; Dobson, C. M.; Heremans, K., High hydrostatic pressure dissociates early aggregates of TTR105-115, but not the mature amyloid fibrils. *Journal of Molecular Biology* **2005,** *347* (5), 903-909.
32. Gras, S. L.; Tickler, A. K.; Squires, A. M.; Devlin, G. L.; Horton, M. A.; Dobson, C. M.; MacPhee, C. E., Functionalised amyloid fibrils for roles in cell adhesion. *Biomaterials* **2008,** *29* (11), 1553-1562.
33. Gazit, E., The "Correctly folded" state of proteins: Is it a metastable state. *Angewandte Chemie-International Edition* **2002,** *41* (2), 257-+.
34. Carny, O.; Gazit, E., A model for the role of short self-assembled peptides in the very early stages of the origin of life. *Faseb Journal* **2005,** *19* (9), 1051-1055.
35. Kardos, J.; Yamamoto, K.; Hasegawa, K.; Naiki, H.; Goto, Y., Direct measurement of the thermodynamic parameters of amyloid formation by isothermal titration calorimetry. *Journal of Biological Chemistry* **2004,** *279* (53), 55308-55314.
36. Caughey, B.; Lansbury, P. T., Protofibrils, pores, fibrils, and neurodegeneration: Separating the responsible protein aggregates from the innocent bystanders. *Annual Review of Neuroscience* **2003,** *26*, 267-298.
37. Sousa, M. M.; Cardoso, I.; Fernandes, R.; Guimaraes, A.; Saraiva, M. J., Deposition of transthyretin in early stages of familial amyloidotic polyneuropathy - Evidence for toxicity of nonfibrillar aggregates. *American Journal of Pathology* **2001,** *159* (6), 1993-2000.
38. Reixach, N.; Deechongkit, S.; Jiang, X.; Kelly, J. W.; Buxbaum, J. N., Tissue damage in the amyloidoses: Transthyretin monomers and nonnative oligomers are the major cytotoxic species in tissue culture. *Proceedings of the National Academy of Sciences of the United States of America* **2004,** *101* (9), 2817-2822.
39. Bucciantini, M.; Giannoni, E.; Chiti, F.; Baroni, F.; Formigli, L.; Zurdo, J. S.; Taddei, N.; Ramponi, G.; Dobson, C. M.; Stefani, M., Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* **2002,** *416* (6880), 507-511.
40. Saraiva, M. J. M., Transthyretin mutations in hyperthyroxinemia and amyloid diseases. *Human Mutation* **2001,** *17* (6), 493-503.
41. Johnson, S. M.; Wiseman, R. L.; Sekijima, Y.; Green, N. S.; Adamski-Werner, S. L.; Kelly, J. W., Native state kinetic stabilization as a strategy to ameliorate protein misfolding diseases: A focus on the transthyretin amyloidoses. *Accounts of Chemical Research* **2005,** *38* (12), 911-921.
42. Damas, A. M.; Saraiva, M. J., Review: TTR amyloidosis - Structural features leading to protein aggregation and their implications on therapeutic strategies. *Journal of Structural Biology* **2000,** *130* (2-3), 290-299.
43. Harper, J. D.; Wong, S. S.; Lieber, C. M.; Lansbury, P. T., Assembly of A beta amyloid protofibrils: An in vitro model for a possible early event in Alzheimer's disease. *Biochemistry* **1999,** *38* (28), 8972-8980.
44. Harper, J. D.; Lieber, C. M.; Lansbury, P. T., Atomic force microscopic imaging of seeded fibril formation and fibril branching by the Alzheimer's disease amyloid-beta protein. *Chemistry & Biology* **1997,** *4* (12), 951-959.
45. Harper, J. D.; Wong, S. S.; Lieber, C. M.; Lansbury, P. T., Observation of metastable A beta amyloid protofibrils by atomic force microscopy. *Chemistry & Biology* **1997,** *4* (2), 119-125.
46. Garai, K.; Sahoo, B.; Sengupta, P.; Maiti, S., Quasihomogeneous nucleation of amyloid beta yields numerical bounds for the critical radius, the surface tension, and the free energy barrier for nucleus formation. *Journal of Chemical Physics* **2008,** *128* (4); Tjernberg, L. O.; Pramanik, A.; Bjorling, S.; Thyberg, P.; Thyberg, J.; Nordstedt, C.; Berndt, K. D.; Terenius, L.; Rigler, R., Amyloid beta-peptide polymerization studied using fluorescence correlation spectroscopy. *Chemistry & Biology* **1999,** *6* (1), 53-62.
47. Roher, A. E.; Chaney, M. O.; Kuo, Y. M.; Webster, S. D.; Stine, W. B.; Haverkamp, L. J.; Woods, A. S.; Cotter, R. J.; Tuohy, J. M.; Krafft, G. A.; Bonnell, B. S.; Emmerling, M. R., Morphology

and toxicity of A beta-(1-42) dimer derived from neuritic and vascular amyloid deposits of Alzheimer's disease. *Journal of Biological Chemistry* **1996,** *271* (34), 20631-20635.
48. Bader, R.; Bamford, R.; Zurdo, J.; Luisi, B. F.; Dobson, C. M., Probing the mechanism of amyloidogenesis through a tandem repeat of the PI3-SH3 domain suggests a generic model for protein aggregation and fibril formation. *Journal of Molecular Biology* **2006,** *356* (1), 189-208.
49. Kayed, R.; Head, E.; Thompson, J. L.; McIntire, T. M.; Milton, S. C.; Cotman, C. W.; Glabe, C. G., Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* **2003,** *300* (5618), 486-489; Kayed, R.; Head, E.; Sarsoza, F.; Saing, T.; Cotman, C. W.; Necula, M.; Margol, L.; Wu, J.; Breydo, L.; Thompson, J. L.; Rasool, S.; Gurlo, T.; Butler, P.; Glabe, C. G., Fibril specific, conformation dependent antibodies recognize a generic epitope common to amyloid fibrils and fibrillar oligomers that is absent in prefibrillar oligomers. *Molecular Neurodegeneration* **2007,** *2*.
50. Williams, A. D.; Sega, M.; Chen, M. L.; Kheterpal, I.; Geva, M.; Berthelier, V.; Kaleta, D. T.; Cook, K. D.; Wetzel, R., Structural properties of A beta protofibrils stabilized by a small molecule. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** *102* (20), 7115-7120.
51. Kheterpal, I.; Zhou, S.; Cook, K. D.; Wetzel, R., A beta amyloid fibrils possess a core structure highly resistant to hydrogen exchange. *Proceedings of the National Academy of Sciences of the United States of America* **2000,** *97* (25), 13597-13601.
52. Levine, H., SOLUBLE MULTIMERIC ALZHEIMER BETA(1-40) PRE-AMYLOID COMPLEXES IN DILUTE-SOLUTION. *Neurobiology of Aging* **1995,** *16* (5), 755-764.
53. Bitan, G.; Lomakin, A.; Teplow, D. B., Amyloid beta-protein oligomerization - Prenucleation interactions revealed by photo-induced cross-linking of unmodified proteins. *Journal of Biological Chemistry* **2001,** *276* (37), 35176-35184.
54. Bitan, G.; Teplow, D. B., Rapid photochemical cross-linking - A new tool for studies of metastable, amyloidogenic protein assemblies. *Accounts of Chemical Research* **2004,** *37* (6), 357-364.
55. Smith, A. M.; Jahn, T. R.; Ashcroft, A. E.; Radford, S. E., Direct observation of oligomeric species formed in the early stages of amyloid fibril formation using electrospray ionisation mass spectrometry. *Journal of Molecular Biology* **2006,** *364* (1), 9-19.
56. Walsh, D. M.; Lomakin, A.; Benedek, G. B.; Condron, M. M.; Teplow, D. B., Amyloid beta-protein fibrillogenesis - Detection of a protofibrillar intermediate. *Journal of Biological Chemistry* **1997,** *272* (35), 22364-22372.
57. Walsh, D. M.; Hartley, D. M.; Kusumoto, Y.; Fezoui, Y.; Condron, M. M.; Lomakin, A.; Benedek, G. B.; Selkoe, D. J.; Teplow, D. B., Amyloid beta-protein fibrillogenesis - Structure and biological activity of protofibrillar intermediates. *Journal of Biological Chemistry* **1999,** *274* (36), 25945-25952.
58. Hou, X.; Aguilar, M. I.; Small, D. H., Transthyretin and familial amyloidotic polyneuropathy - Recent progress in understanding the molecular mechanism of neurodegeneration. *Febs Journal* **2007,** *274* (7), 1637-1650.
59. Carulla, N.; Caddy, G. L.; Hall, D. R.; Zurdo, J.; Gairi, M.; Feliz, M.; Giralt, E.; Robinson, C. V.; Dobson, C. M., Molecular recycling within amyloid fibrils. *Nature* **2005,** *436* (7050), 554-558.
60. Deng, W.; Cao, A. N.; Lai, L. H., Detecting the inter-peptide arrangement and maturation process of transthyretin (105-115) amyloid fibril using a FRET pair with short Forster distance. *Biochemical and Biophysical Research Communications* **2007,** *362* (3), 689-694.
61. Magill, J. H., Review spherulites: A personal perspective. *Journal of Materials Science* **2001,** *36* (13), 3143-3164.
62. Krebs, M. R. H.; Bromley, E. H. C.; Rogers, S. S.; Donald, A. M., The mechanism of amyloid spherulite formation by bovine insulin. *Biophysical Journal* **2005,** *88* (3), 2013-2021.
63. Krebs, M. R. H.; Devlin, G. L.; Donald, A. M., Protein particulates: Another generic form of protein aggregation? *Biophysical Journal* **2007,** *92* (4), 1336-1342.
64. Krebs, M. R. H.; MacPhee, C. E.; Miller, A. F.; Dunlop, L. E.; Dobson, C. M.; Donald, A. M., The formation of spherulites by amyloid fibrils of bovine insulin. *Proceedings of the National Academy of Sciences of the United States of America* **2004,** *101* (40), 14420-14424; Rogers, S. S.; Krebs, M. R. H.; Bromley, E. H. C.; van der Linden, E.; Donald, A. M., Optical microscopy of growing insulin amyloid spherulites on surfaces in vitro. *Biophysical Journal* **2006,** *90* (3), 1043-1054; Ruth, L.; Eisenberg, D.;

Neufeld, E. F., alpha-L-iduronidase forms semi-crystalline spherulites with amyloid-like properties. *Acta Crystallographica Section D-Biological Crystallography* **2000,** *56*, 524-528.
65. Aggeli, A.; Bell, M.; Carrick, L. M.; Fishwick, C. W. G.; Harding, R.; Mawer, P. J.; Radford, S. E.; Strong, A. E.; Boden, N., pH as a trigger of peptide beta-sheet self-assembly and reversible switching between nematic and isotropic phases. *Journal of the American Chemical Society* **2003,** *125* (32), 9619-9628.
66. Sobott, F.; Hernandez, H.; McCammon, M. G.; Tito, M. A.; Robinson, C. V., A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Analytical Chemistry* **2002,** *74* (6), 1402-1407.
67. Teplow, D. B.; Lazo, N. D.; Bitan, G.; Bernstein, S.; Wyttenbach, T.; Bowers, M. T.; Baumketner, A.; Shea, J. E.; Urbanc, B.; Cruz, L.; Borreguero, J.; Stanley, H. E., Elucidating amyloid beta-protein folding and assembly: A multidisciplinary approach. *Accounts of Chemical Research* **2006,** *39* (9), 635-645.
68. Bernstein, S. L.; Liu, D. F.; Wyttenbach, T.; Bowers, M. T.; Lee, J. C.; Gray, H. B.; Winkler, J. R., alpha-synuclein: Stable compact and extended monomeric structures and pH dependence of dimer formation. *Journal of the American Society for Mass Spectrometry* **2004,** *15* (10), 1435-1443.
69. Bowers, M. T., Conformation and Aggregation of Biological Molecules: The Latest News. In *Isolated Biomolecules and Biomolecular Interactions*, Valladolid, Spain, 2008.
70. Bernstein, S. L.; Wyttenbach, T.; Baumketner, A.; Shea, J. E.; Bitan, G.; Teplow, D. B.; Bowers, M. T., Amyloid beta-protein: Monomer structure and early aggregation states of A beta 42 and its Pro(19) alloform. *Journal of the American Chemical Society* **2005,** *127* (7), 2075-2084.
71. Bitan, G.; Kirkitadze, M. D.; Lomakin, A.; Vollers, S. S.; Benedek, G. B.; Teplow, D. B., Amyloid beta-protein (A beta) assembly: A beta 40 and A beta 42 oligomerize through distinct pathways. *Proceedings of the National Academy of Sciences of the United States of America* **2003,** *100* (1), 330-335.
72. Baumketner, A.; Bernstein, S. L.; Wyttenbach, T.; Bitan, G.; Teplow, D. B.; Bowers, M. T.; Shea, J. E., Amyloid beta-protein monomer structure: A computational and experimental study. *Protein Science* **2006,** *15* (3), 420-428.
73. Zhang, S.; Iwata, K.; Lachenmann, M. J.; Peng, J. W.; Li, S.; Stimson, E. R.; Lu, Y.; Felix, A. M.; Maggio, J. E.; Lee, J. P., The Alzheimer's peptide A beta adopts a collapsed coil structure in water. *Journal of Structural Biology* **2000,** *130* (2-3), 130-141.
74. Borysik, A. J. H.; Radford, S. E.; Ashcroft, A. E., Co-populated conformational ensembles of beta(2)-microglobulin uncovered quantitatively by electrospray ionization mass spectrometry. *Journal of Biological Chemistry* **2004,** *279* (26), 27069-27077.
75. Borysik, A. J. H.; Read, P.; Little, D. R.; Bateman, R. H.; Radford, S. E.; Ashcroft, A. E., Separation of beta(2)-microglobulin conformers by high-field asymmetric waveform ion mobility spectrometry (FAIMS) coupled to electrospray ionisation mass spectrometry. *Rapid Communications in Mass Spectrometry* **2004,** *18* (19), 2229-2234.
76. Myers, S. L.; Thomson, N. H.; Radford, S. E.; Ashcroft, A. E., Investigating the structural properties of amyloid-like fibrils formed in vitro from beta 2-microglobulin using limited proteolysis and electrospray ionisation mass spectrometry. *Rapid Communications in Mass Spectrometry* **2006,** *20* (11), 1628-1636.
77. Sobott, F.; McCammon, M. G.; Robinson, C. V., Gas-phase dissociation pathways of a tetrameric protein complex. *International Journal of Mass Spectrometry* **2003,** *230* (2-3), 193-200.
78. Wysocki, V. H.; Joyce, K. E.; Jones, C. M.; Beardsley, R. L., Surface-induced dissociation of small molecules, peptides,and non-covalent protein complexes. *Journal of the American Society for Mass Spectrometry* **2008,** *19* (2), 190-208; Wells, J. M.; McLuckey, S. A., Collision-induced dissociation (CID) of peptides and proteins. In *Biological Mass Spectrometry*, 2005; Vol. 402, pp 148-185.
79. Ruotolo, B. T.; Hyung, S. J.; Robinson, P. M.; Giles, K.; Bateman, R. H.; Robinson, C. V., Ion mobility-mass spectrometry reveals long-lived, unfolded intermediates in the dissociation of protein complexes. *Angewandte Chemie-International Edition* **2007,** *46* (42), 8001-8004.
80. Nettleton, E. J.; Sunde, M.; Lai, Z. H.; Kelly, J. W.; Dobson, C. M.; Robinson, C. V., Protein subunit interactions and structural integrity of amyloidogenic transthyretins: Evidence from electrospray mass spectrometry. *Journal of Molecular Biology* **1998,** *281* (3), 553-564.

81. Nettleton, E. J.; Tito, P.; Sunde, M.; Bouchard, M.; Dobson, C. M.; Robinson, C. V., Characterization of the oligomeric states of insulin in self-assembly and amyloid fibril formation by mass spectrometry. *Biophysical Journal* **2000,** *79* (2), 1053-1065; Tito, P.; Nettleton, E. J.; Robinson, C. V., Dissecting the hydrogen exchange properties of insulin under amyloid fibril forming conditions: A site-specific investigation by mass spectrometry. *Journal of Molecular Biology* **2000,** *303* (2), 267-278.
82. Knapman, T. W.; Aggeli, A.; Ashcroft, A. E., Critical concentrations of beta-sheet peptide self-assembly quantified directly by nanoelectrospray ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* **2008,** *22* (10), 1611-1614.
83. Clemmer, D. E.; Jarrold, M. F., Ion mobility measurements and their applications to clusters and biomolecules. *Journal of Mass Spectrometry* **1997,** *32* (6), 577-592; Jarrold, M. F., Peptides and proteins in the vapor phase. *Annual Review of Physical Chemistry* **2000,** *51*, 179-207.
84. MacPhee, C., 2008.
85. Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F., Structural information from ion mobility measurements: Effects of the long-range potential. *Journal of Physical Chemistry* **1996,** *100* (40), 16082-16086.
86. Li, D. W.; Han, L.; Huo, S. H., Structural and pathway complexity of beta-strand reorganization within aggregates of human transthyretin(105-115) peptide. *Journal of Physical Chemistry B* **2007,** *111* (19), 5425-5433.
87. Murray, M. M.; Bernstein, S. L.; Nyugen, V.; Condron, M. M.; Teplow, D. B.; Bowers, M. T., Amyloid beta Protein: A beta 40 Inhibits A beta 42 Oligomerization. *Journal of the American Chemical Society* **2009,** *131* (18), 6316-+.
88. Krone, M. G.; Baumketner, A.; Bernstein, S. L.; Wyttenbach, T.; Lazo, N. D.; Teplow, D. B.; Bowers, M. T.; Shea, J. E., Effects of familial Alzheimer's disease mutations on the folding nucleation of the amyloid beta-protein. *Journal of Molecular Biology* **2008,** *381* (1), 221-228.

6

Conclusions

This thesis contains results from the application of a combined theoretical and experimental methodology to the assessment of polypeptide and peptide aggregate conformation and conformational stability in a solvent-free environment. The two case-studies presented (Trp cage, TTR105-115) agree with the consensus view of protein folding studied in a gas-phase environment[1].

Trp cage is in essence a synthetic protein folding model. A question that persists among protein research by MS can be expressed as "what is the smallest protein whose conformation can be transferred in the gas phase?" IMMS data presented here offer quite compelling evidence for stating that Trp cage, despite it being stabilised in solution by both strong electrostatics and hydrophobic interactions, undergoes significant structural changes upon desolvation. Thus a search for the smallest protein whose conformation persists the electrospray process continues. However, results obtained for Trp cage imply that it is not simply system size that affects gas-phase stability but the nature of interactions it contains. Structures formed by strong electrostatic interactions (hydrogen bonding, ion-dipole interactions) appear to survive in the gas phase (this is essentially the same conclusion reached by Patriksson *et al.* from vacuum REM simulations and ECD fragmentation pathways of Trp cage constructs[2]).

Another trait that is shared among systems studied is the compact configuration they adopt in the absence of solvent. Structural contraction of proteins in the gas phase became apparent even from early protein IMMS experiments (see for example Clemmer

and Jarrold[3]). However, not only are peptide ions containing few charges compact, but these gas-phase conformations possess a remarkable thermal stability. These results imply that intrinsically stable 'non-native' compact configurations exist in isolated peptide systems. In the presence of a polar solvent such structures are, presumably, entropically disfavoured as such rigid conformations reduce the number of H-bonds that can form with water molecules; however their existence and isolation in the gas phase illuminates the nature of 'unfolded' protein states. Therefore, a denatured protein ensemble may not occupy exclusively extended configurations.

The thermal stability of protein ions, as exemplified by Trp cage IMMS experiments, has further implications for the applicability of MS-based techniques for studying biomolecular conformations. Due to the presence of several like charges in ESI-generated ions, which become solvated by the rest of the protein molecule, energy barriers to conformational transitions are likely to be increased. Gas-phase techniques may present a very viable alternative for isolating, separating and characterising distinct non-covalent arrangements which co-exist in solution. Peptide folding may be studied in great detail using MS-based methods; these possibilities have already begun being exploited[4].

In addition to single macromolecule folding, the investigation of aggregating peptide systems is another "difficult to address" problem. However, due to the high mass resolution of modern mass spectrometers, the isolation and characterisation of such macromolecular assemblies in the gas phase is possible. IMMS and molecular modelling have been applied to interrogate the structures of early-appearing oligomers of TTR105-115, a system of interest in the design of new self-assembling materials as well as a model for studying amyloid fibril formation. A series of oligomers (from dimer to dodecamer) of TTR105-115 was detected by MS; ion mobility measurements revealed that the structures of these oligomers are globular. Furthermore, simulations of oligomeric states up to the hexamer, as well as experimental data, seem to imply, even in this early stage of the project, an important role for the tetrameric aggregate. These findings are of some interest especially considering the similarities in amyloidogenesis mechanisms for different systems. Should the results described in Chapter 5 be

corroborated by further research (currently under way in Dr. Barran's group in collaboration with Dr. MacPhee) it would appear that further large-scale conformational rearrangements are required to generate fibrils from these 'intermediate' oligomeric states.

The gas phase offers a minimal and controllable environment in which the intrinsic stability of molecular assemblies can be studied. With the development for novel instrumentation and ongoing computational progress, a greater understanding of the principles of polypeptide structure will result from such investigations.

References

1. Clemmer, D. E.; Jarrold, M. F., Ion mobility measurements and their applications to clusters and biomolecules. *Journal of Mass Spectrometry* **1997,** *32* (6), 577-592; Hoaglund-Hyzer, C. S.; Counterman, A. E.; Clemmer, D. E., Anhydrous protein ions. *Chemical Reviews* **1999,** *99* (10), 3037-3079; Jarrold, M. F., Peptides and proteins in the vapor phase. *Annual Review of Physical Chemistry* **2000,** *51*, 179-207; Wyttenbach, T.; Bowers, M. T., Intermolecular interactions in biomolecular systems examined by mass spectrometry. *Annual Review of Physical Chemistry* **2007,** *58*, 511-533.

2. Patriksson, A.; Adams, C.; Kjeldsen, F.; Raber, J.; van der Spoel, D.; Zubarev, R. A., Prediction of N-C-alpha bond cleavage frequencies in electron capture dissociation of Trp-cage dications by force-field molecular dynamics simulations. *International Journal of Mass Spectrometry* **2006,** *248* (3), 124-135; Patriksson, A.; Adams, C. M.; Kjeldsen, F.; Zubarev, R. A.; van der Spoel, D., A direct comparison of protein structure in the gas and solution phase: The TRP-cage. *Journal of Physical Chemistry B* **2007,** *111*, 13147-13150.

3. Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F., NAKED PROTEIN CONFORMATIONS - CYTOCHROME-C IN THE GAS-PHASE. *Journal of the American Chemical Society* **1995,** *117* (40), 10141-10142.

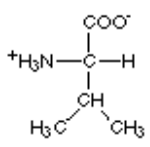
4. Bohrer, B. C.; Mererbloom, S. I.; Koeniger, S. L.; Hilderbrand, A. E.; Clemmer, D. E., Biomolecule Analysis by Ion Mobility Spectrometry. *Annual Review of Analytical Chemistry* **2008,** *1*, 293-327.

Appendix A

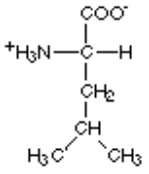
Nomenclature, Structure and Gas-phase Basicity of 20 Naturally Occurring Amino Acids

Name	Three-letter code	One-letter code	Gas-phase basicity[1] kcal mol^{-1}
Glycine	Gly	G	202.7
Alanine	Ala	A	206.4
Serine	Ser	S	207.6
Threonine	Thr	T	211.7
Cysteine	Cys	C	206.2
Methionine	Met	M	213.3
Tyrosine	Tyr	Y	213.1
Tryptophan	Trp	W	216.1
Phenylalanine	Phe	F	212.1
Valine	Val	V	208.7
Leucine	Leu	L	209.6
Isoleucine	Ile	I	210.8
Proline	Pro	P	214.3
Asparagine	Asn	N	212.8
Glutamine	Gln	Q	214.2
Aspartic acid	Asp	D	208.6
Glutamic acid	Glu	E	215.6
Histidine	His	H	223.7
Lysine	Lys	K	221.8
Arginine	Arg	R	237

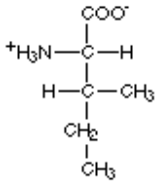
Amino acids with hydrophobic side groups



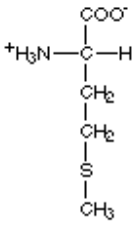
Valine
(val)



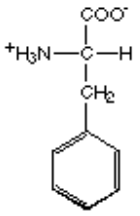
Leucine
(leu)



Isoleucine
(ile)

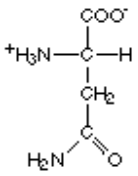


Methionine
(met)

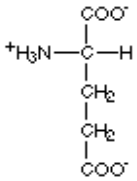


Phenylalanine
(phe)

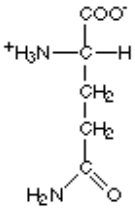
Amino acids with hydrophilic side groups



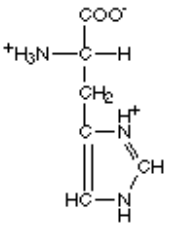
Asparagine
(asn)



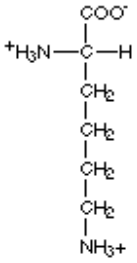
Glutamic acid
(glu)



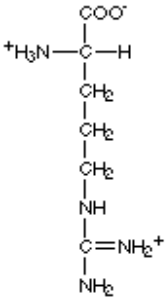
Glutamine
(gln)



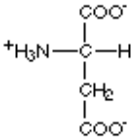
Histidine
(his)



Lysine
(lys)

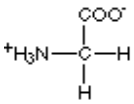


Arginine
(arg)

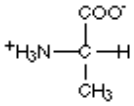


Aspartic acid
(asp)

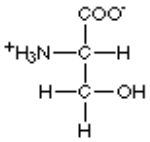
Amino acids that are in between



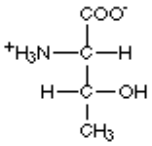
Glycine
(gly)



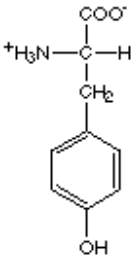
Alanine
(ala)



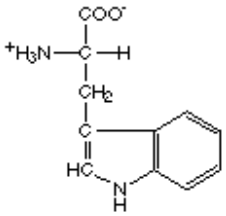
Serine
(ser)



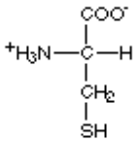
Threonine
(thr)



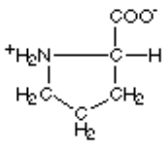
Tyrosine
(tyr)



Tryptophan
(trp)



Cysteine
(cys)



Proline
(pro)

1. Harrison, A. G., The gas-phase basicities and proton affinities of amino acids and peptides. *Mass Spectrometry Reviews* **1997,** *16* (4), 201-217.

Appendix B

Fundamentals of ion transport in drift-tube ion mobility experiments

Ion mobility experiments assume a central role throughout this thesis. An overview of elementary derivations of the most important relationships essential for carrying out this work provide a useful introduction for the method employed and outline the theoretical framework on the basis of which IMMS experiments are interpreted, despite the fact that these relations are well-presented elsewhere along with a more detailed discussion[2, 3].

We will separate this introduction into four sections, each focusing on a different relationship describing ion transport. These are:

1. Statement of the *equation of continuity – the transport equation*, define the differential equations which govern drift and diffusion of ions in a weak field. Solution of the transport equation for the one-dimensional case.
2. Derivation of the *Nernst-Einstein-Townsend relation*, linking the diffusion coefficient to the ionic mobility.
3. Description of the drift velocity and the ionic mobility using simple arguments of momentum conservation, valid for systems undergoing elastic collisions.

B.1 The transport equation

Ions in drift tubes (under linear, non-oscillating and low fields) undergo two processes; drift and diffusion. Therefore Fick's first law can be expressed as

$$\mathbf{J} = \mathbf{v}_D n - D\nabla n \qquad \{B.1.1\}$$

and the continuity equation takes the form[3]

$$\frac{\partial n}{\partial t} = -\nabla \mathbf{J} = D\nabla^2 n - \mathbf{v}_D \nabla n \qquad \{B.1.2\}$$

In order to allow depleting reactions to occur (with a reaction rate constant k_-) a term must be added at the right-hand side of the above expression:

$$\frac{\partial n}{\partial t} = -\nabla \mathbf{J} = D\nabla^2 n - \mathbf{v}_D \nabla n - k\_n \qquad \{B.1.3\}$$

We shall now proceed to solve {B.1.3} for the 1-D case of ions being initially concentrated at one point at the entrance of the drift-tube and drifting along the long axis of the drift tube. In one dimension:

$$\frac{\partial n}{\partial t} = D\frac{\partial^2 n}{\partial x^2} - v_D \frac{\partial n}{\partial x} - k\_n \qquad \{B.1.4\}$$

The ion density n may be represented with its fourier transform over the space dimension. Thus, in reciprocal space (k) {B.1.4} becomes:

$$\frac{\partial \tilde{n}(k,t)}{\partial t} = (ik)^2 D\tilde{n}(k,t) - ikv_D\tilde{n}(k,t) - k\_\tilde{n}(k,t) \qquad \{B.1.5\}$$

Collecting terms and solving the resulting 1st order DE:

$$\tilde{n}(k,t) = \tilde{n}(k,0)e^{-\left(k^2 D + iv_D k + k\_\right)t} \qquad \{B.1.6\}$$

To obtain the ion density in real space, the inverse transform $n(x.t) = \frac{1}{2\pi}\int \tilde{n}(k,t)e^{ikx}dk$ must be taken. The initial condition $n(x,0)$ can be given by a delta function:

$$n(x,0) = a\delta(x) \qquad \{B.1.7\}$$

whose F.T. can be computed

$$\tilde{n}(k,0) = a\int \delta(x-0)e^{-ikx}dx = a \qquad \{B.1.8\}$$

hence,

$$n(x,t) = \frac{a}{2\pi}\int e^{ikx-\left(Dk^2+iv_D k+k\_\right)t}dk =$$

$$\frac{ae^{-k\_t}}{2\pi}\int e^{-Dt\left(k^2-\frac{i(x-v_D t)}{Dt}-\left(\frac{x-v_D t}{2Dt}\right)^2+\left(\frac{x-v_D t}{2Dt}\right)^2\right)}dk =$$

$$\frac{ae^{-\frac{(x-v_D t)^2}{4Dt}-k\_t}}{2\pi}\int e^{-Dt\left(k-\frac{i(x-v_D t)}{2Dt}\right)^2}dk = \frac{ae^{-\frac{(x-v_D t)^2}{4Dt}-k\_t}}{2\pi}\sqrt{\frac{\pi}{Dt}} = \qquad \{B.1.9\}$$

$$\frac{ae^{-\frac{(x-v_D t)^2}{4Dt}-k\_t}}{\sqrt{4\pi Dt}}$$

The axial ion current can be estimated from {B.1.2}. Differentiating {B.1.9} and rearranging gives the axial ion current:

$$J(x,t) = \frac{a}{4\sqrt{\pi D t}}\left(v_D + \frac{x}{t}\right)e^{-\frac{(x-v_D t)^2}{4Dt} - k\_t} \qquad \{B.1.10\}$$

Thus ATDs from IMS experiments can be fitted to an equation like {B.1.10} assuming that the remainder of the instrument does not distort the ion distribution significantly. Analysis for different initial conditions is given by Barnes[4]; further analysis using a Green's function method is discussed by Gatland[5], who also reaches a formulation that takes into account inter-converting species.

B.2 Mobility and Diffusion at vanishing fields: the Nernst-Einstein-Townsend equation

The dependence of ion speed to the force applied via the mobility is a fundamental result for IMMS experiments – it has been stated repeatedly in chapters 1 and 2 and is discussed further the following section. On the other hand, phenomena pertaining to diffusion seem to be mediated by an empirically established diffusion coefficient, *D*. The expressions for the ion flux and the current (equations {B.1.9} & {B.1.10}) imply that drift and diffusion are concurrent but independent processes occurring during drift-tube experiments. The Nernst-Einstein-Townsend equation relates the two processes when the system is found near equilibrium, i.e. in the case of vanishingly weak fields.

At equilibrium, the spatial distribution of ions *n* can be represented by a Boltzmann exponential:

$$n = n_0 e^{\frac{q\mathbf{E}\bullet\mathbf{r}}{k_B T}} \qquad \{B.2.1\}$$

The gradient of *n* is therefore given by

$$\nabla n = \frac{q\mathbf{E}}{k_B T} n_0 e^{\frac{q\mathbf{E}\bullet\mathbf{r}}{k_B T}} = \frac{q\mathbf{E}}{k_B T} n \qquad \{B.2.2\}$$

Inserting this expression in equation {B.1.1}, setting $\mathbf{J} = 0$ (meaning there is no net ion flux at equilibrium) and rearranging we obtain

$$n\mathbf{E}\left(K-\frac{qD}{k_BT}\right)=0\Rightarrow$$
$$K=\frac{qD}{k_BT} \qquad \{B.2.3\}$$

This result is known as the Nernst-Einstein-Townsend relation and shows that close to the zero-field limit drift and diffusion phenomena are interrelated. Written as $\mathbf{v}_D = D\frac{q\mathbf{E}}{k_BT}$ the relation clarifies the fact that at vanishing fields ionic velocity is proportional to the ratio of the Lorenz force and average thermal energy, the diffusion coefficient being the constant of proportionality between the two.

B.3 Description of ionic drift velocity based on conservation of energy[*]

During an IMS experiment, ions experience two forces. Due to their charge they tend to accelerate along the direction of the field applied (electrical component of the Lorentz force):

$$\mathbf{F}_{\mathrm{L}} = q\mathbf{E}\Rightarrow \mathbf{v}_{i,vac}(t)=\frac{q\mathbf{E}t}{m_i} \quad \{B.3.1\}$$

(**F**: force; *q*: charge; **E**: electric field; **v**: velocity; *t*: time; *m*: mass; subscript *i* indicates ions). However, as collisions with neutrals (in our case He atoms) are frequent, kinetic energy is redistributed, with the resulting into a macroscopic constant velocity. For a constant drift velocity to result the Lorentz force must be balanced exactly by a friction-like force pointing at the opposite direction; thus the velocity gained due to acceleration between collisions will *on average* be transferred to the neutrals. Defining τ as the average time between consecutive collisions, one may state:

$$\left\langle \mathbf{v'}_n - \mathbf{v}_n \right\rangle = \frac{q\mathbf{E}\tau}{m_n} \quad \{B.3.2\}$$

[*] This derivation follows the one published by Ravercomb and Mason[2]

$\mathbf{v}_n$ is the velocity of neutrals before and $\mathbf{v'}_n$ after a collision; the subscript n represents 'neutrals'. As neutrals undergo random motions, their average velocity $\langle \mathbf{v}_n \rangle$ is zero. Hence:

$$\langle \mathbf{v'}_n \rangle = \frac{q\mathbf{E}\tau}{m_n} \qquad \{B.3.3\}$$

Ion-neutral collisions not only redistribute the energy between the two species, but also the direction of motion. Assuming a nearly spherical geometry for ions and neutrals this randomisation is complete, leading to a zero relative velocity after a collision; *i.e.* average ion and neutral velocities immediately after a collision are equal:

$$\langle \mathbf{v'}_i \rangle = \langle \mathbf{v'}_n \rangle \quad \{B.3.4\}$$

Momentum balance requires that

$$m_i \langle \mathbf{v}_i \rangle + m_n \langle \mathbf{v}_n \rangle = m_i \langle \mathbf{v'}_i \rangle + m_n \langle \mathbf{v'}_n \rangle \qquad \{B.3.5\}$$

However, as stated above, the random distribution of direction of initial neutral velocities cause $\langle \mathbf{v}_n \rangle = 0$; furthermore, average ion and neutral speeds after a collision are equal and 'average ion velocity' is simply a description of the drift velocity, $\mathbf{v}_D$. Thus, rearranging, {B.3.5} becomes

$$\mathbf{v}_D = \left(\frac{1}{m_i} + \frac{1}{m_n} \right) m_n \langle \mathbf{v'}_n \rangle \quad \{B.3.6\}$$

Combining with {B.3.3} and defining the reduced mass $\frac{1}{\mu} \equiv \frac{1}{m_i} + \frac{1}{m_n}$ one yields

$$\mathbf{v}_D = \frac{q\mathbf{E}\tau}{\mu} \quad \{B.3.7\}$$

Reducing τ to measurable quantities can be done by describing the mean-free-path λ (the average distance traversed by an ion between collisions) in terms of the ion-neutral relative speed and τ:

$$\lambda = \langle v_{rel} \rangle \tau \quad \{B.3.8\}$$

By definition, ions encounter one neutral in time τ; therefore, if ions have an orientationally-averaged cross-sectional area Ω, there exist one neutral molecule in volume $\lambda\Omega$. In these terms, the neutral number density N can be expressed:

$$N = \frac{1}{\langle v_{rel} \rangle \tau \Omega} \Rightarrow \tau = \frac{1}{\langle v_{rel} \rangle N \Omega} \qquad \{B.3.9\}$$

The number density for gases like helium around ambient temperature is sufficiently approximated by the ideal gas law $P = Nk_B T$. Now, an estimate of $\langle v_{rel} \rangle$ can complete the derivation. To achieve this end, we must assume $\langle v_{rel} \rangle^2 \approx \langle v_{rel}^2 \rangle$; subsequently proceed with the calculation

$$\langle v_{rel}^2 \rangle = \langle |\mathbf{v}_i - \mathbf{v}_n|^2 \rangle = \langle v_i^2 \rangle + \langle v_n^2 \rangle - 2\langle \mathbf{v}_i \bullet \mathbf{v}_n \rangle \qquad \{B.3.10\}$$

Since the velocities of ions and neutrals are independent, hence uncorrelated, the last term must vanish. Average thermal energy can account for the mean-square speed of neutral atoms. The same also holds for ions if the field energy is negligibly small compared to thermal energy. Thus:

$$\frac{1}{2} m_n \langle v_n^2 \rangle = \frac{3}{2} k_B T \qquad \{B.3.11a\}$$

$$\frac{1}{2} m_i \langle v_i^2 \rangle = \frac{3}{2} k_B T \qquad \{B.3.11b\}$$

Leading to

$$\langle v_{rel} \rangle^2 \approx \langle v_{rel}^2 \rangle = \frac{3k_B T}{m_i} + \frac{3k_B T}{m_n} \Rightarrow \langle v_{rel} \rangle = \sqrt{\frac{3k_B T}{\mu}} \qquad \{B.3.12\}$$

Combining {B.3.12} and {B.3.9} one obtains

$$\tau = \frac{\mu^{1/2}}{(3k_B T)^{1/2} N \Omega} \qquad \{B.3.13\}$$

which, after inserting into {B.3.7} and rearranging, gives

$$\mathbf{v}_D = K\mathbf{E} = \frac{1}{3^{1/2}} \frac{q\mathbf{E}}{N} \left(\frac{1}{\mu k_B T} \right)^{1/2} \frac{1}{\Omega} \qquad \{B.3.14\}$$

{B.3.14} is essentially identical to the kinetic formula for low fields, employed throughout this thesis for estimating cross sections from measured mobilities:

$$\mathbf{v}_D = \frac{3}{16}\frac{q\mathbf{E}}{N}\left(\frac{2\pi}{\mu k_B T}\right)^{1/2}\frac{1}{\Omega} \qquad \{B.3.15\}$$

differing only in some constants. The relationships between all physical quantities are predicted correctly from simple momentum-transfer theory. Moreover, certain basic conditions for the validity of the formula are revealed during this derivation. Most important of these, codified in expression {B.3.13}, states that for this view of ion transport to hold, thermal energy must dominate the kinetic energy of ions. However in the absence of this criterion modifications can be made to amend the situation. Forces encountered along the field axis will increase at stronger fields; these observations open the gates for the development of multiple-temperature theories. Additionally, the amount of energy deposited on analyte ions at higher fields can be estimated[2, 3]. The resulting relationship is the same as {B.3.14} but replacing the temperature T with an effective temperature T_{eff}, given by

$$T_{eff} = T\left(1+\frac{m_n v_D^2}{3k_B T}\right) \qquad \{B.3.16\}$$

a result which enables the formulation of inequalities that describe low-field behaviour and also indicates that at higher fields dependence on the square of the field strength and the mobility arises.

Yet, in the case of protein IMS the problem that arises at higher fields is associated with a) non-spherical geometries and b) redistribution of energy over both kinetic and conformational modes. The ability of high-energy collisions to affect the structure of a large molecule is a factor which albeit significant, cannot be accounted for by this theory without modifications (for a more accurate description of the phenomena at higher fields should take into account the possibility of inelastic collisions).

In addition, other limitations, such as the assumption of negligible ion-neutral and ion-ion interactions, the need for spherical geometries and the prominence of binary collisions, are also revealed in this calculation. Nevertheless, despite the ample

approximations, momentum transfer theory accurately captures all the physical principles of ion motion at low fields.

B.4 References

2. Revercomb, H. E.; Mason, E. A., THEORY OF PLASMA CHROMATOGRAPHY GASEOUS ELECTROPHORESIS - REVIEW. *Analytical Chemistry* **1975,** *47* (7), 970-983.
3. Mason, E. A.; McDaniel, E. W., *Transport properties of ions in gases*. Wiley: New York, 1988; p xvi, 560 p.
4. Barnes, W. S., METHOD OF ANALYSIS OF ION SWARM EXPERIMENTS. *Physics of Fluids* **1967,** *10* (9P1), 1941-&.
5. R., G. I., Analysis for Ion Drift Tube Experiments. *Case Studies in Atomic Physics* **1974,** *4*, 369-437.

Appendix C

Additional Methodological Information

C.1 Error estimates

Whenever multiple repeats of particular experiments were performed, a confidence interval was calculated according to the following formula:

$$\text{uncertainty}(\alpha, s, n) = F^{-1}(1-\alpha/2)\frac{s}{\sqrt{n}} \qquad \{\text{App.C.1}\}$$

In the above expression s is the standard deviation, n the statistical population (number of replicas), (1-α) the confidence level and $F^{-1}(p)$ the inverse of the cumulative density function of a standard normal distribution (mean of 0 and standard deviation of 1) at probability p. For a 99% confidence interval (α = 0.01, p = 0.995) $F^{-1}(0.995) \approx 2.576$; this interval was chosen for values described herein (NB, this criterion is more stringent than the 2×s interval so oft encountered).

If only single measurements were available, an error estimate of 5% was arbitrarily assumed. This level of uncertainty was found to be in line with the ones obtained by more rigorous error analysis described.

Appendix D

Supplementary Data

D.1 Visual comparison between experimental and simulated annealing collision cross sections

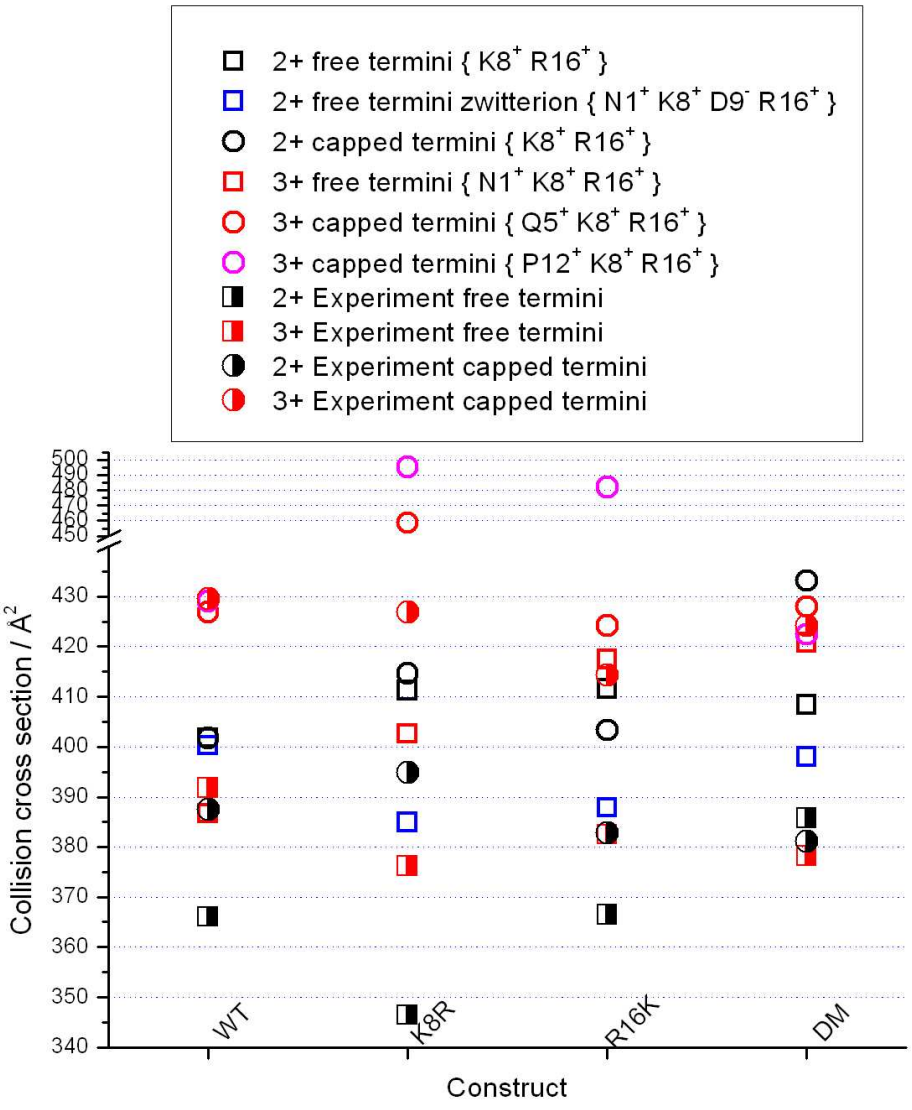


Figure App.D.1 Experimentally-determined collision cross sections of Trp cage ions compared to scaled cross sections from simulated annealing runs for Trp cage constructs. Half-filled symbols: experimental cross sections; open symbols: calculated cross sections.